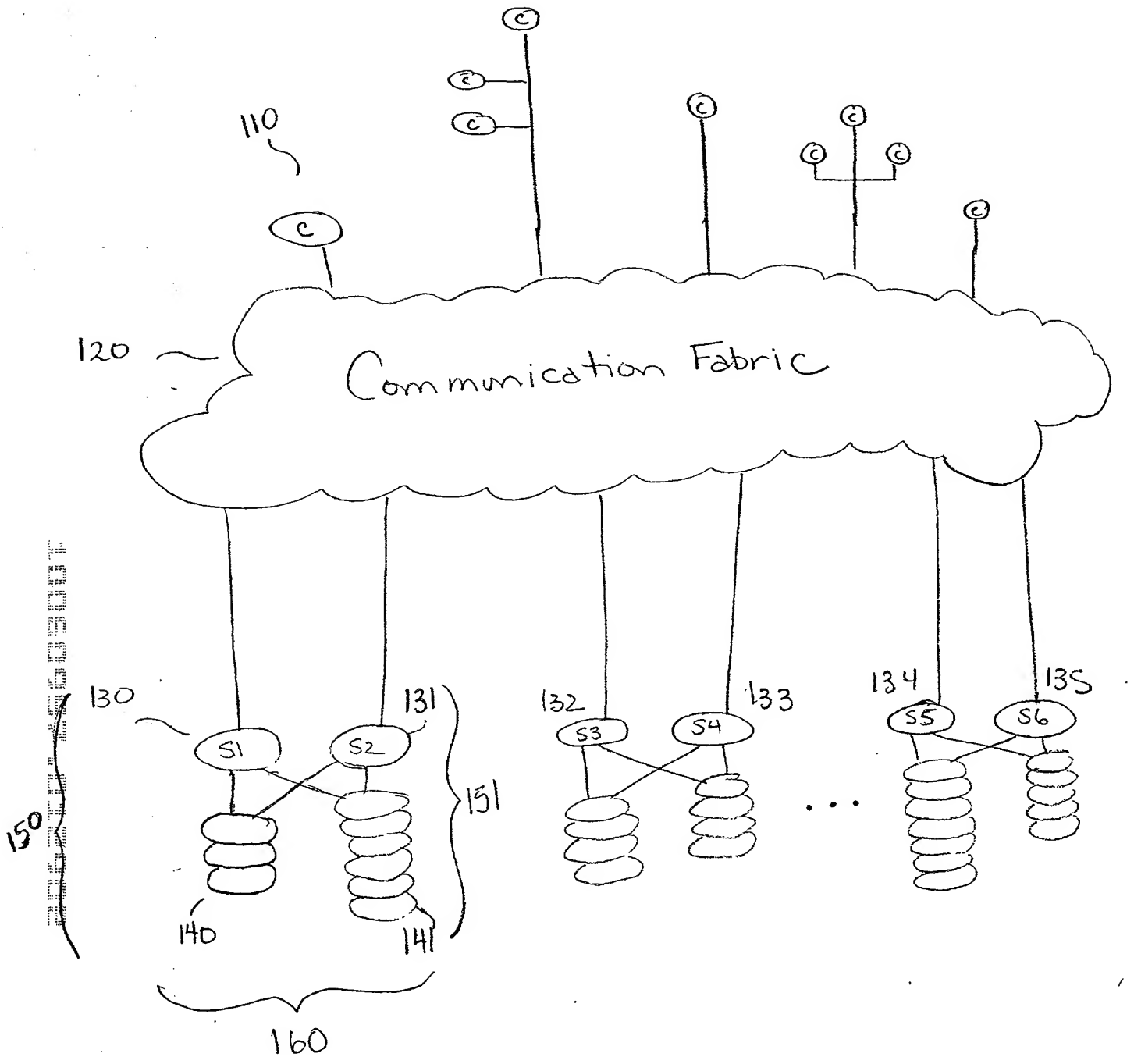


FIGURE 1 - General Overview of Distributed File Storage System

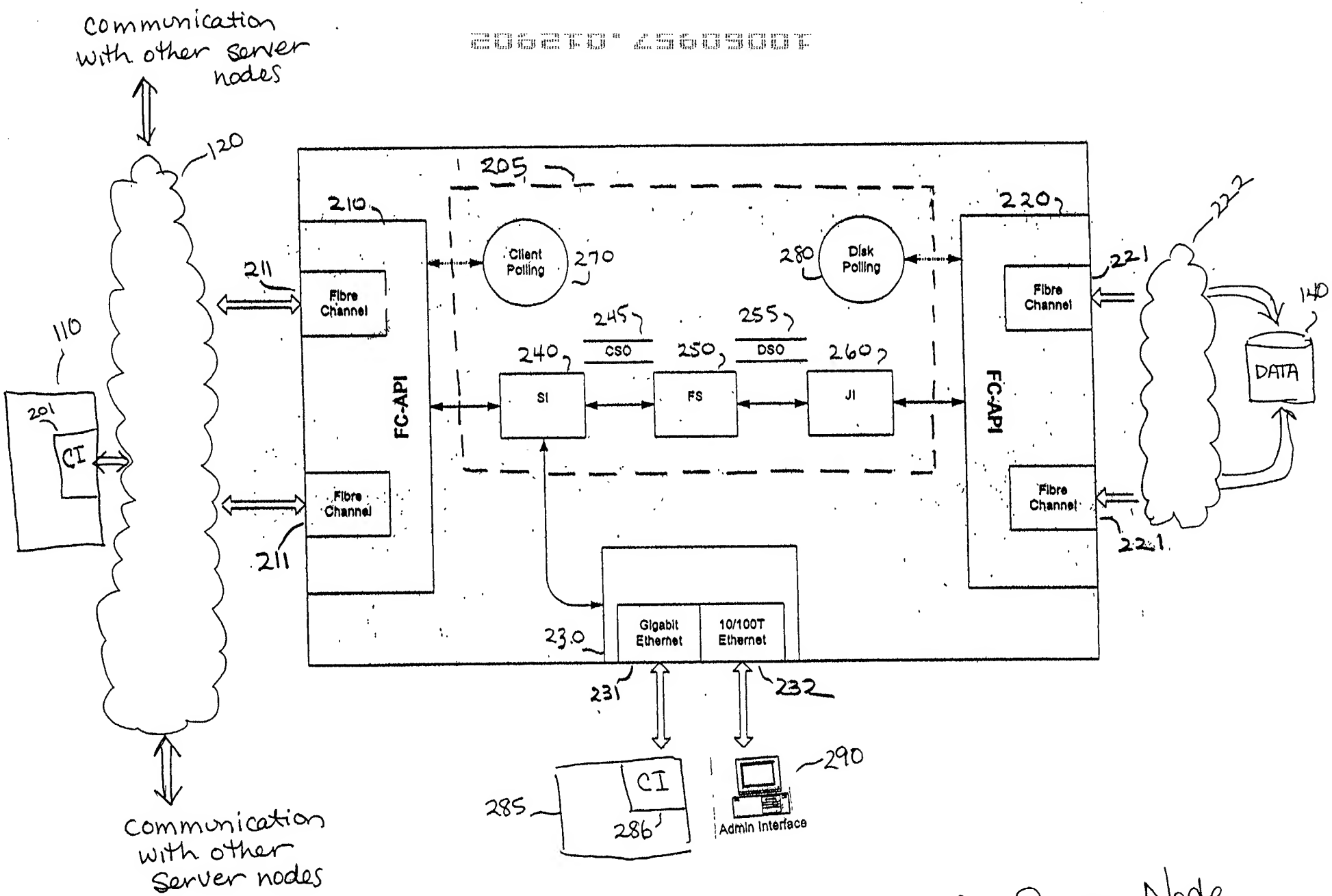


FIGURE 2: One Embodiment of a Server Node

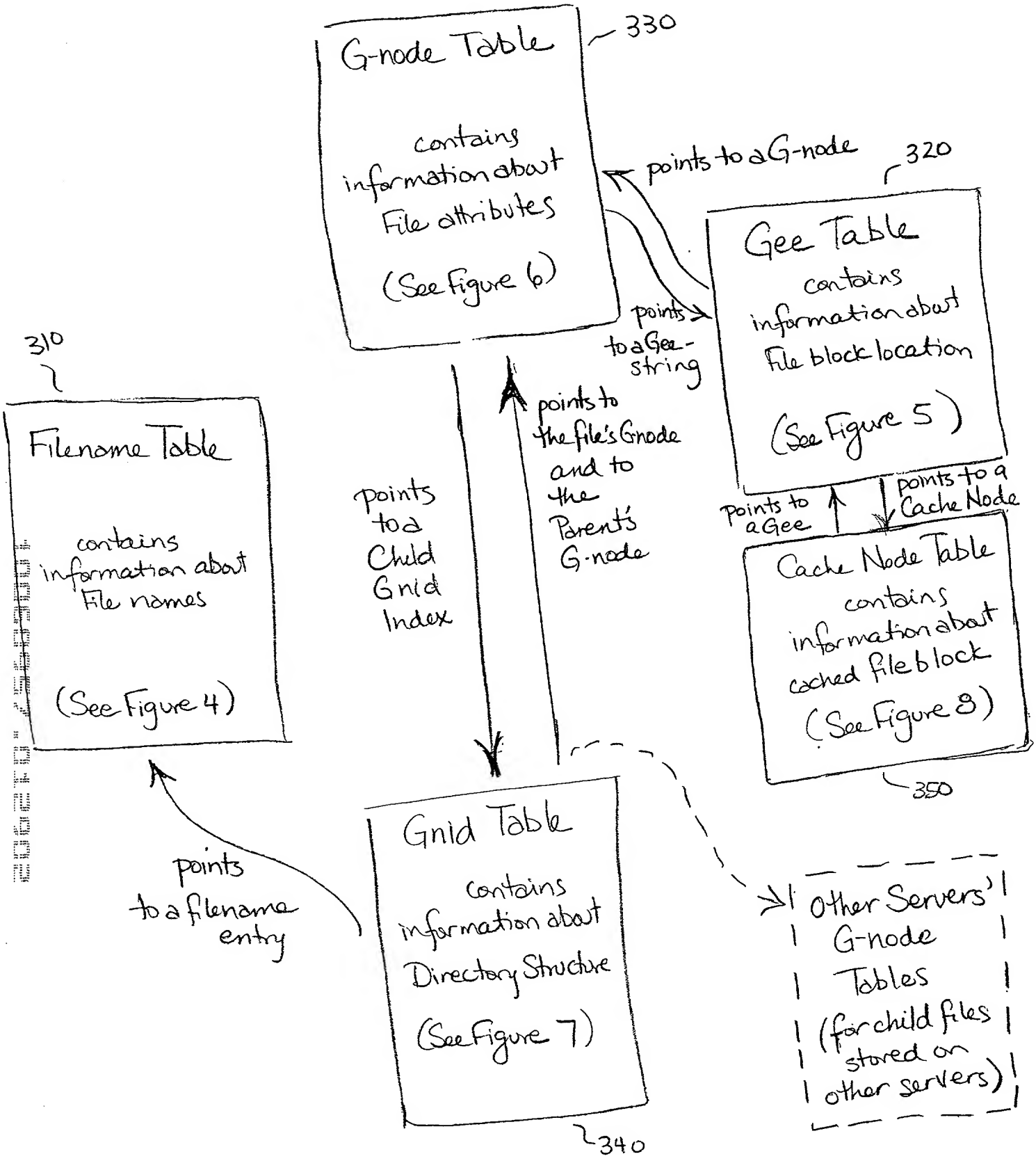


FIGURE 3 - Five metadata structures

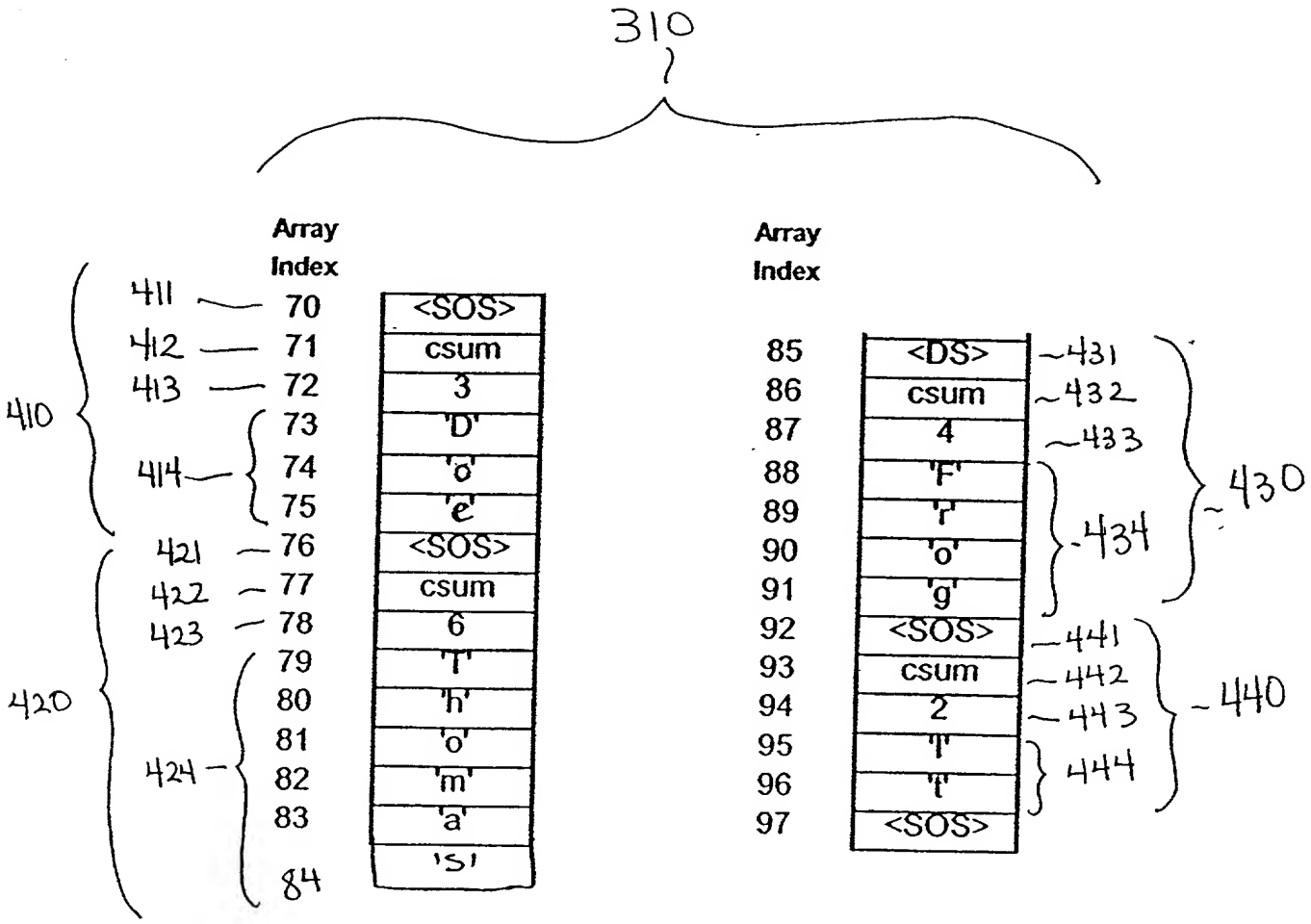


FIGURE 4 - Sample Portion of a Filename Table

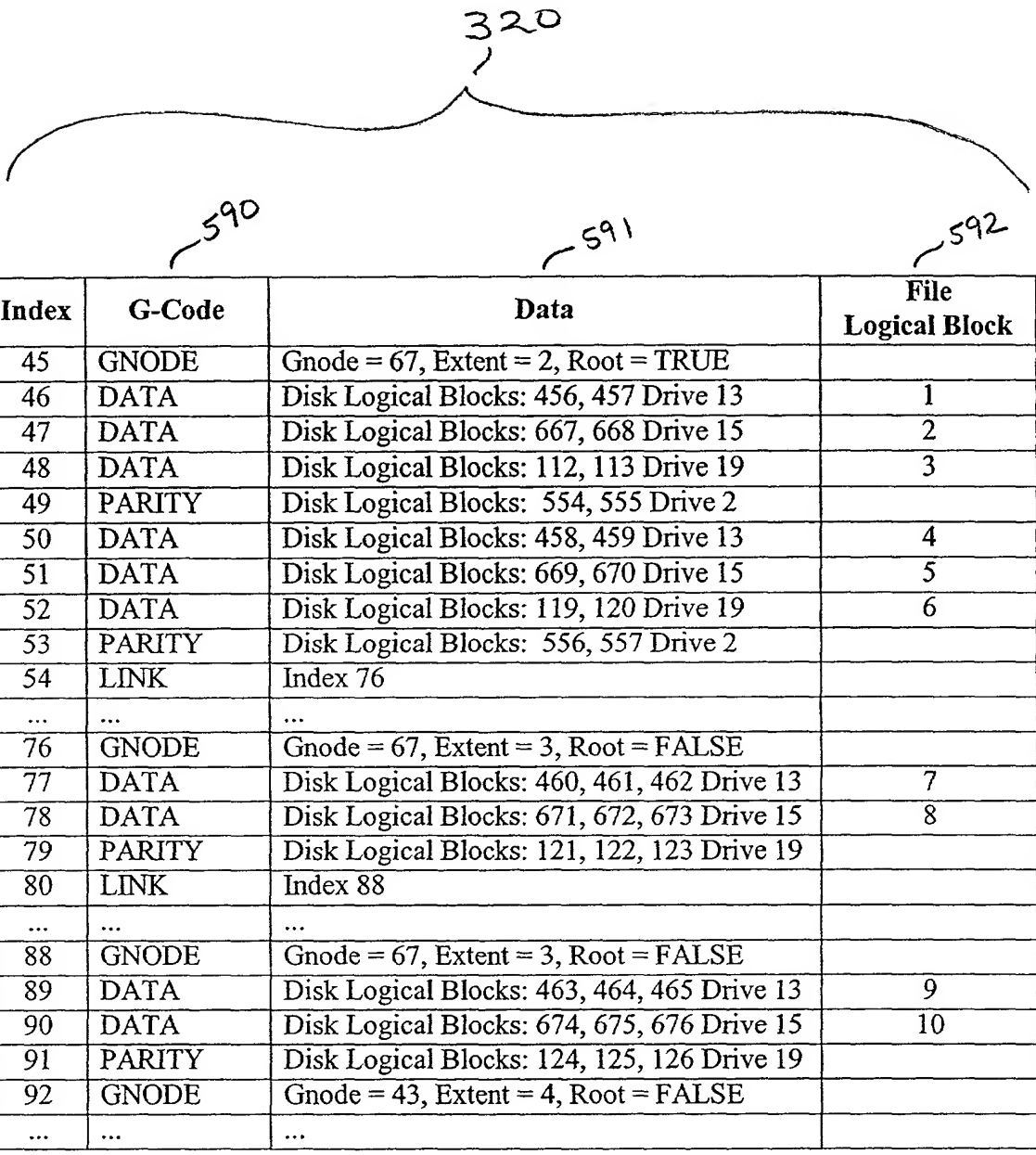


FIGURE 5 - Sample Portion of a Gee Table

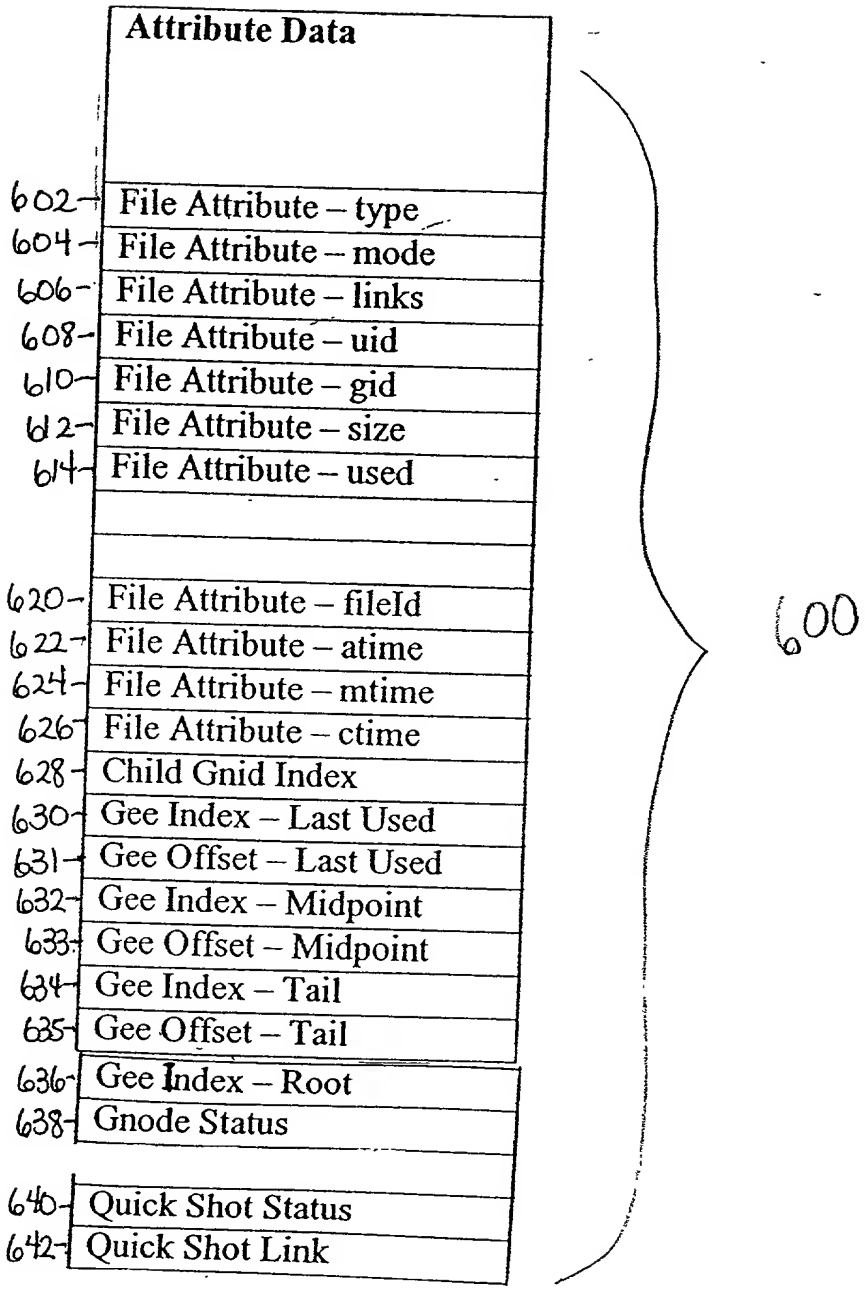


FIGURE 6 - G-NODE ATTRIBUTES

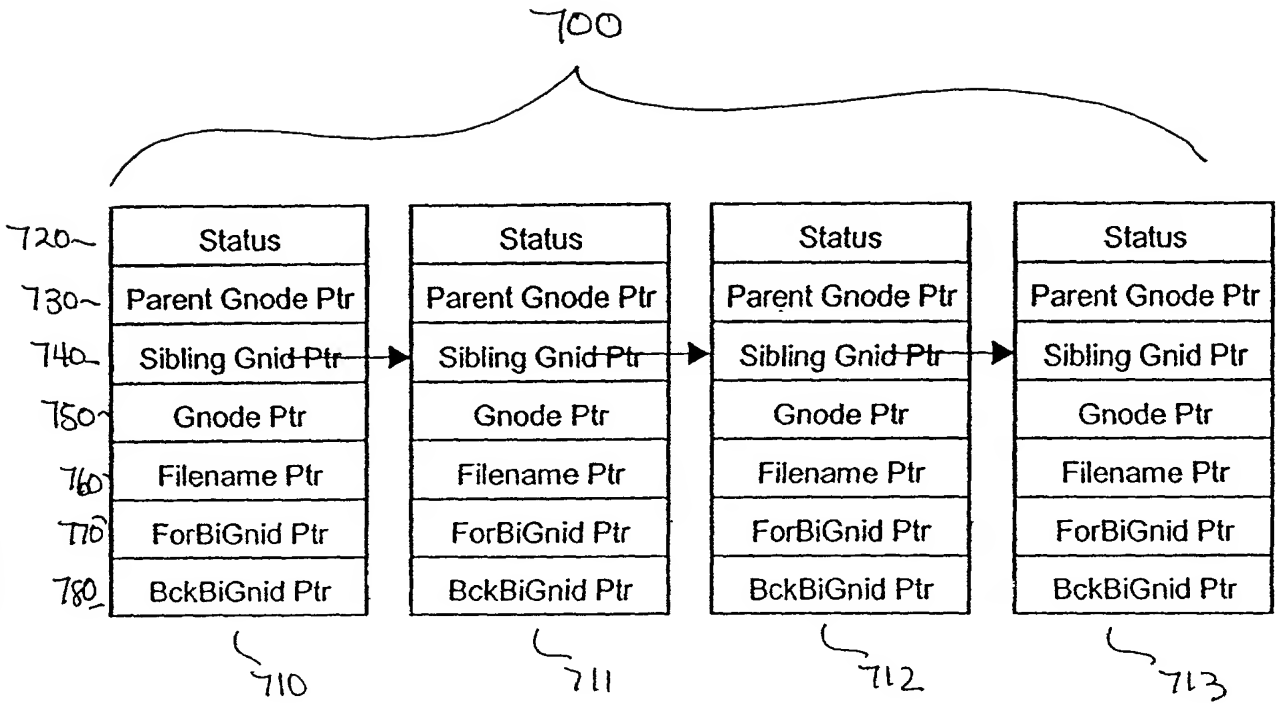


FIGURE 7- Structure of a Gnid String

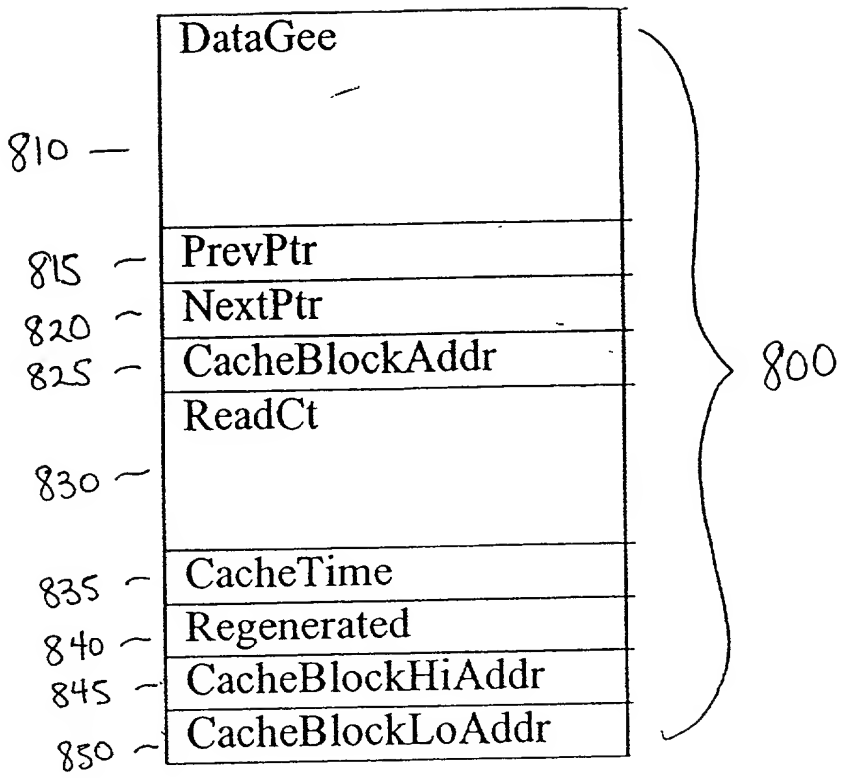


FIGURE 8a - Structure of a Cache Node

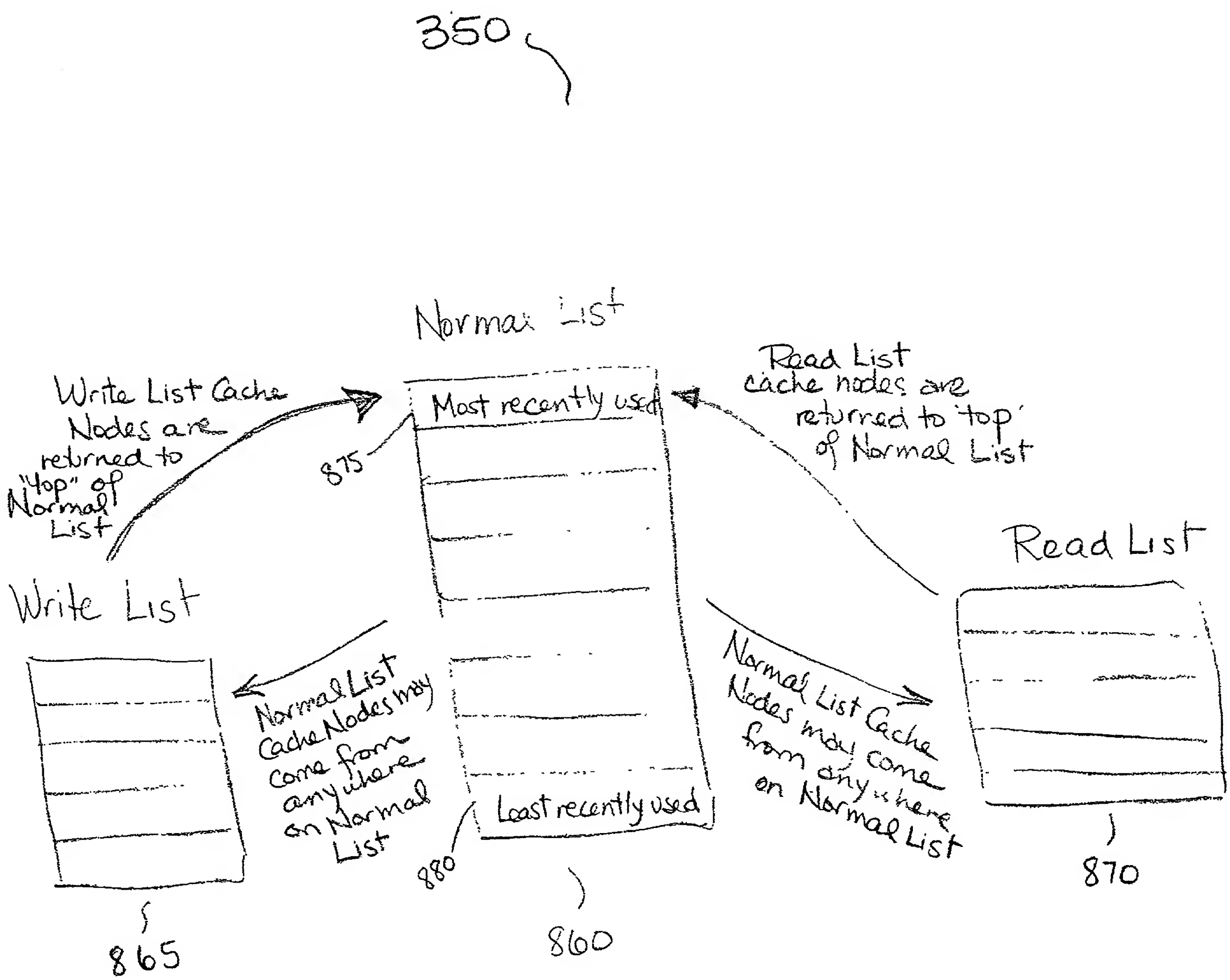


FIGURE 8B - Conceptual division of a Cache Node Table into Three Lists

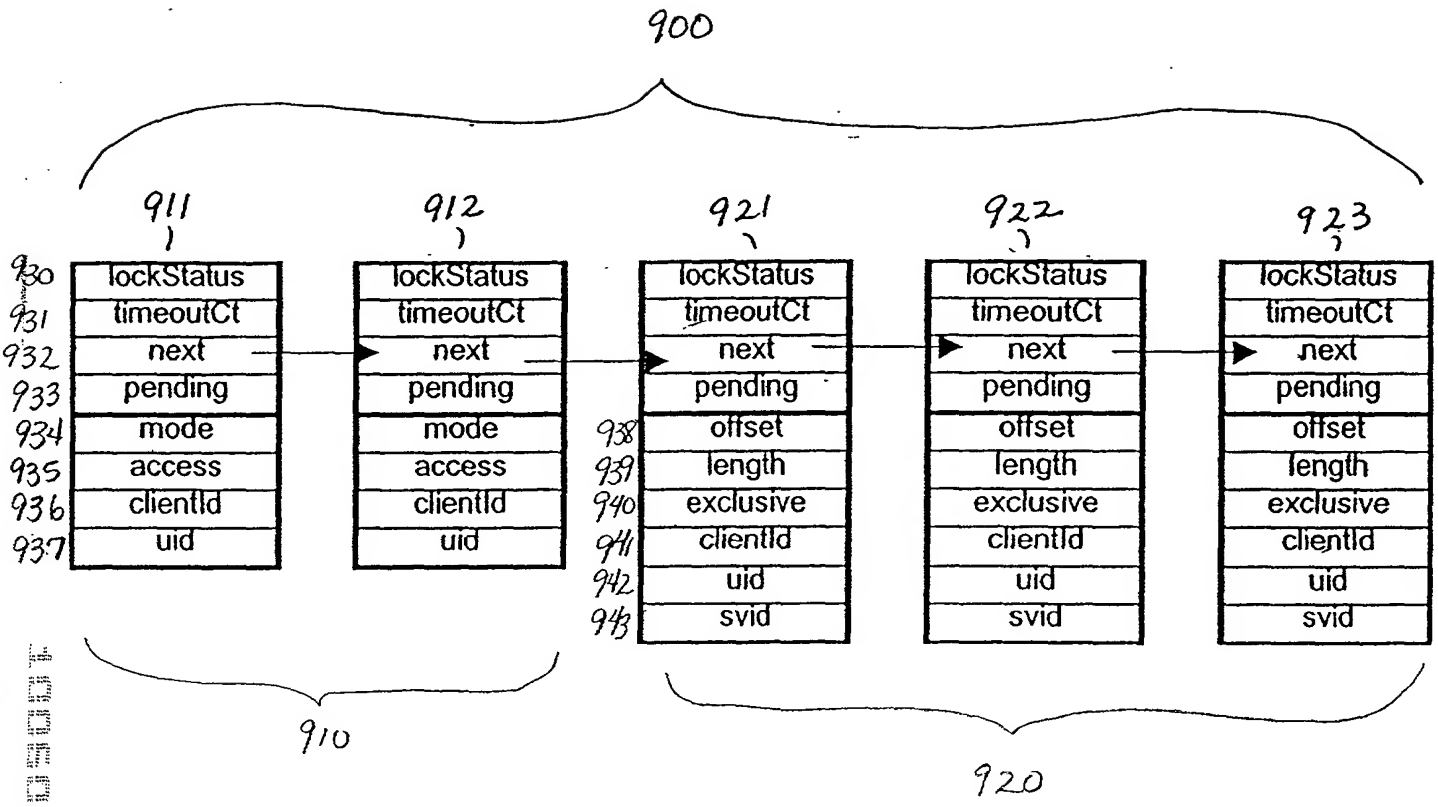


FIGURE 9 - A Sample Lock String

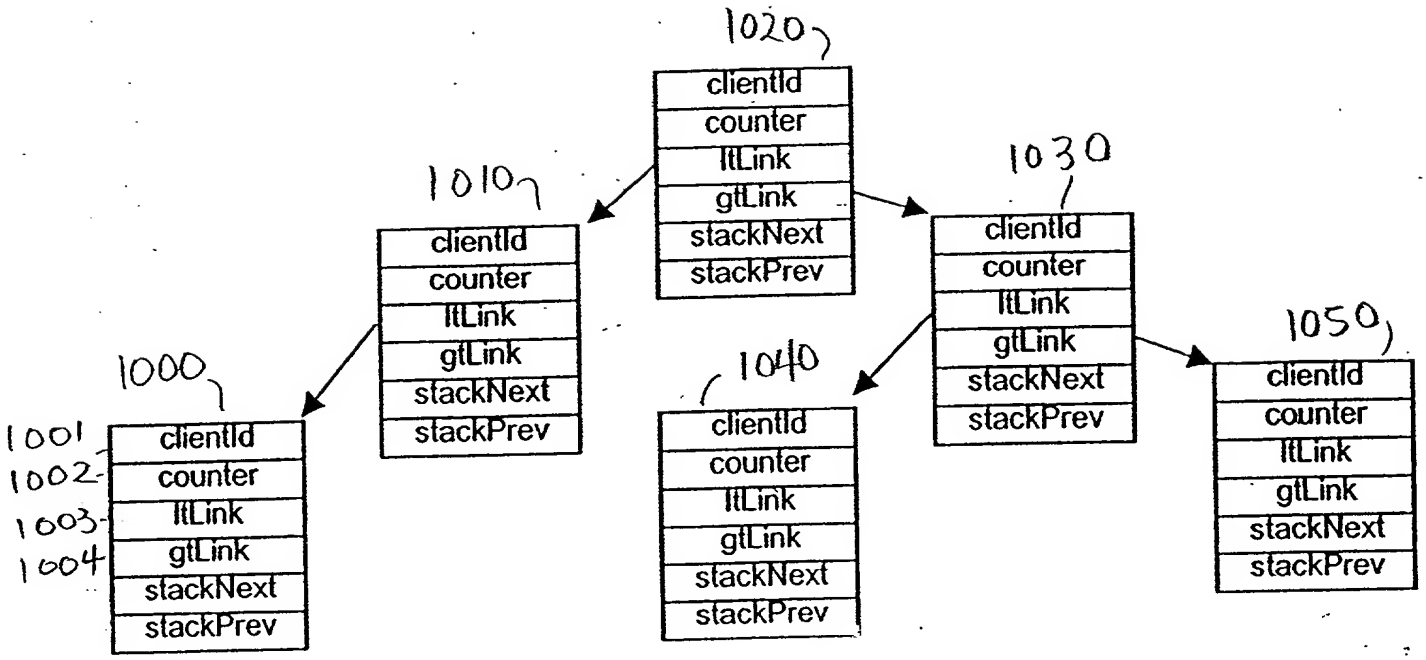


FIGURE 10 - Refresh Nodes configured as a binary tree.

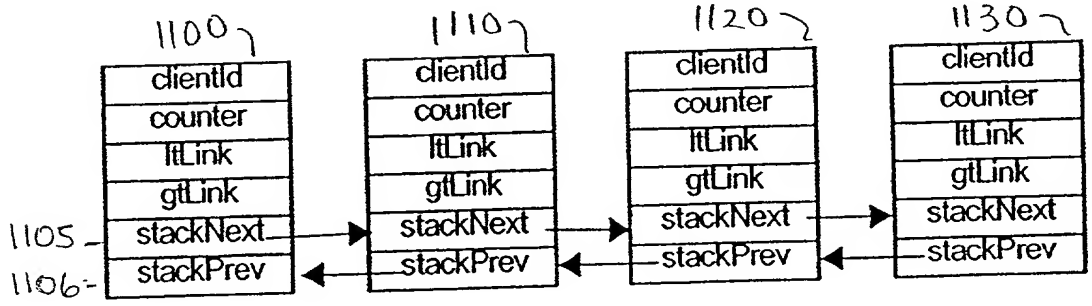


FIGURE 11 - RefreshNodes configured as a doubly-linked list

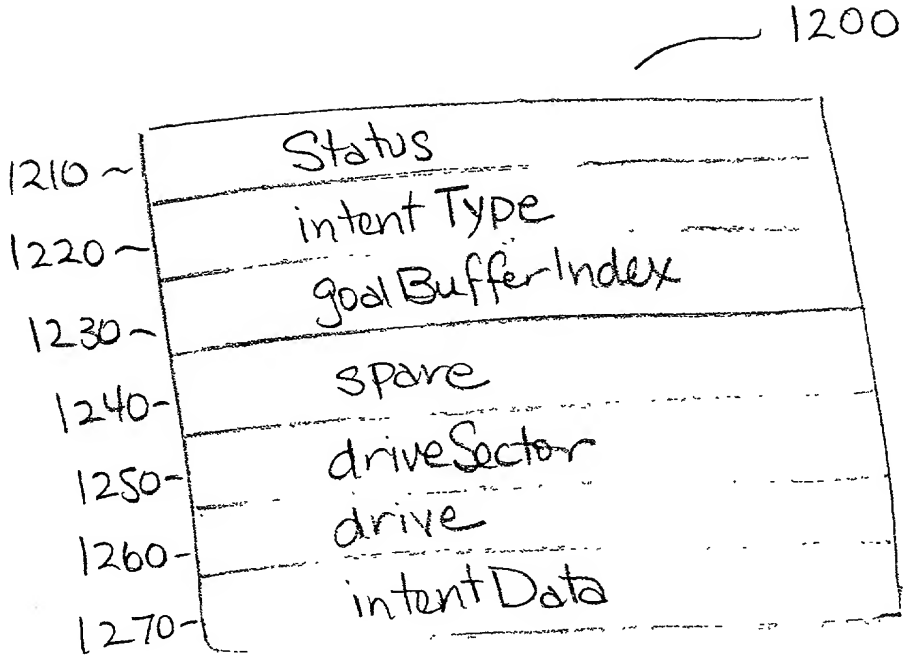


FIGURE 12 - Structure of an Intent Log Entry

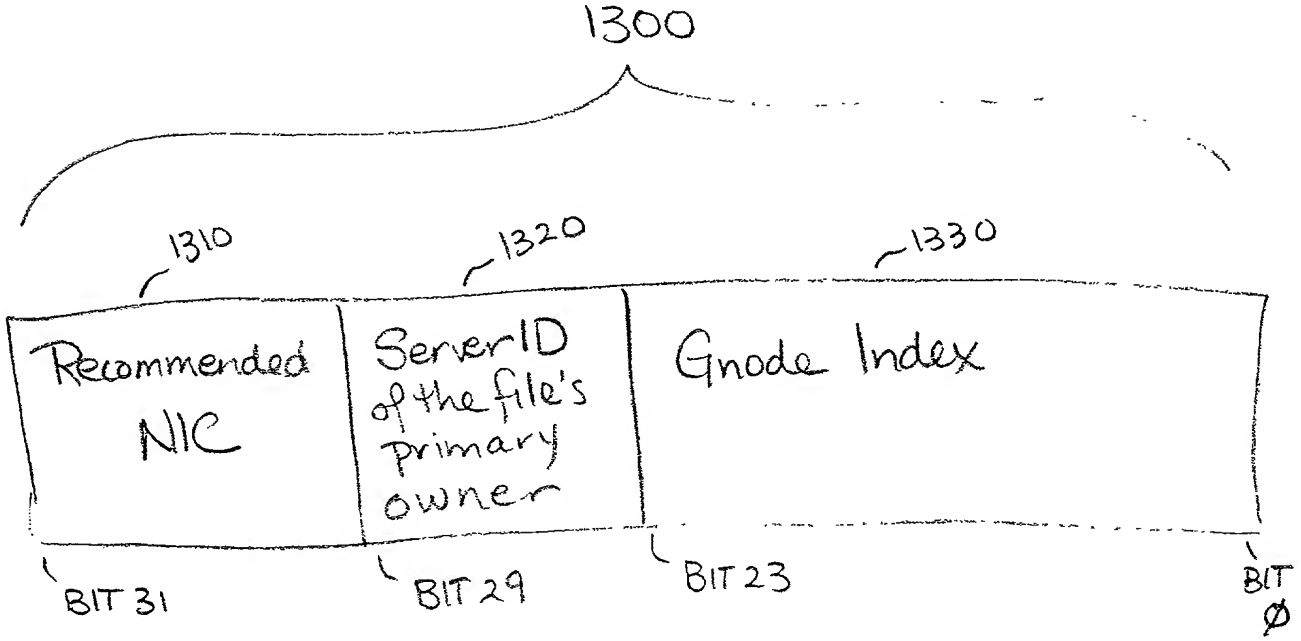


FIGURE 13 - Structure of a File Handle

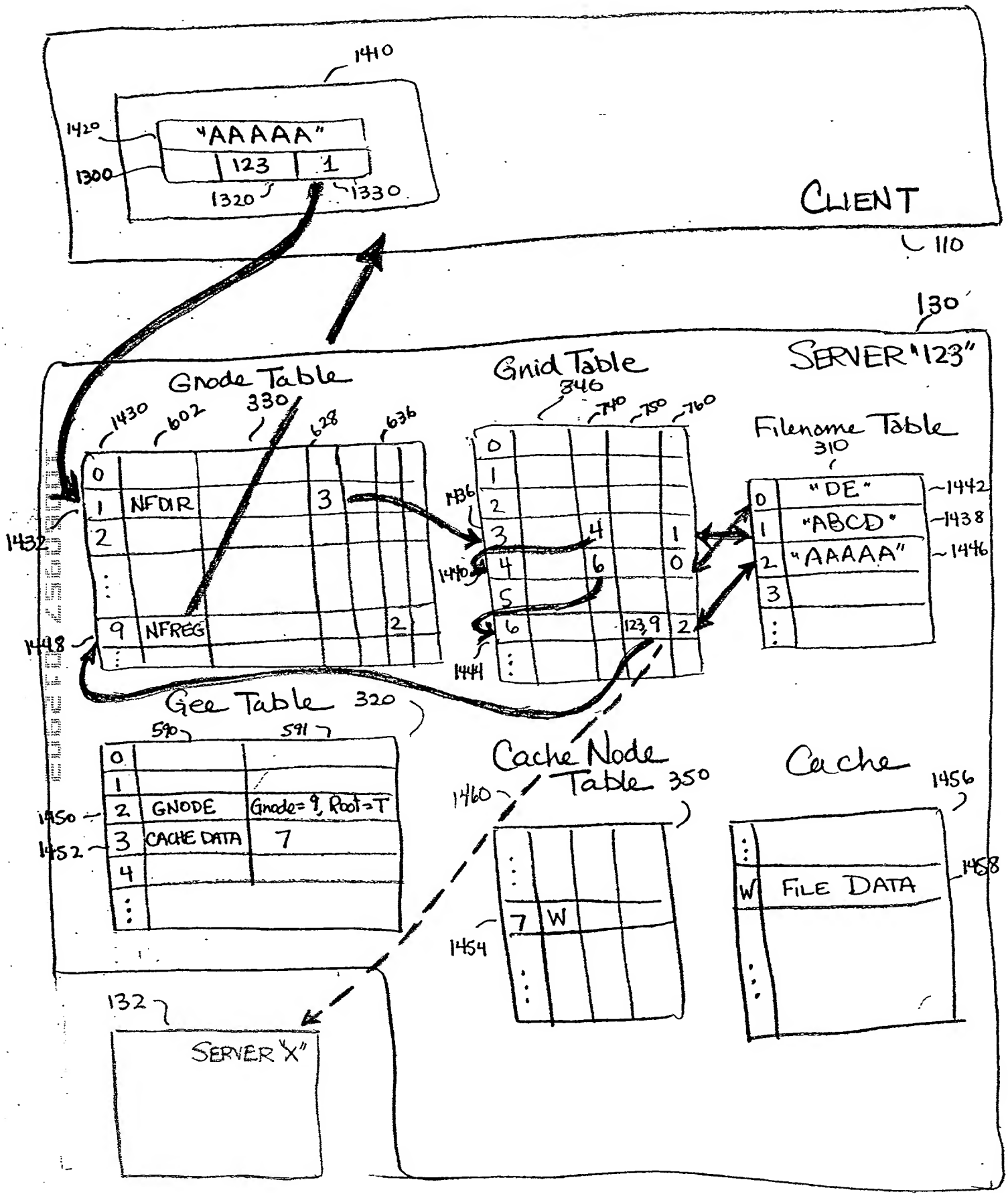


FIGURE 14a: Example of a File Look-Up

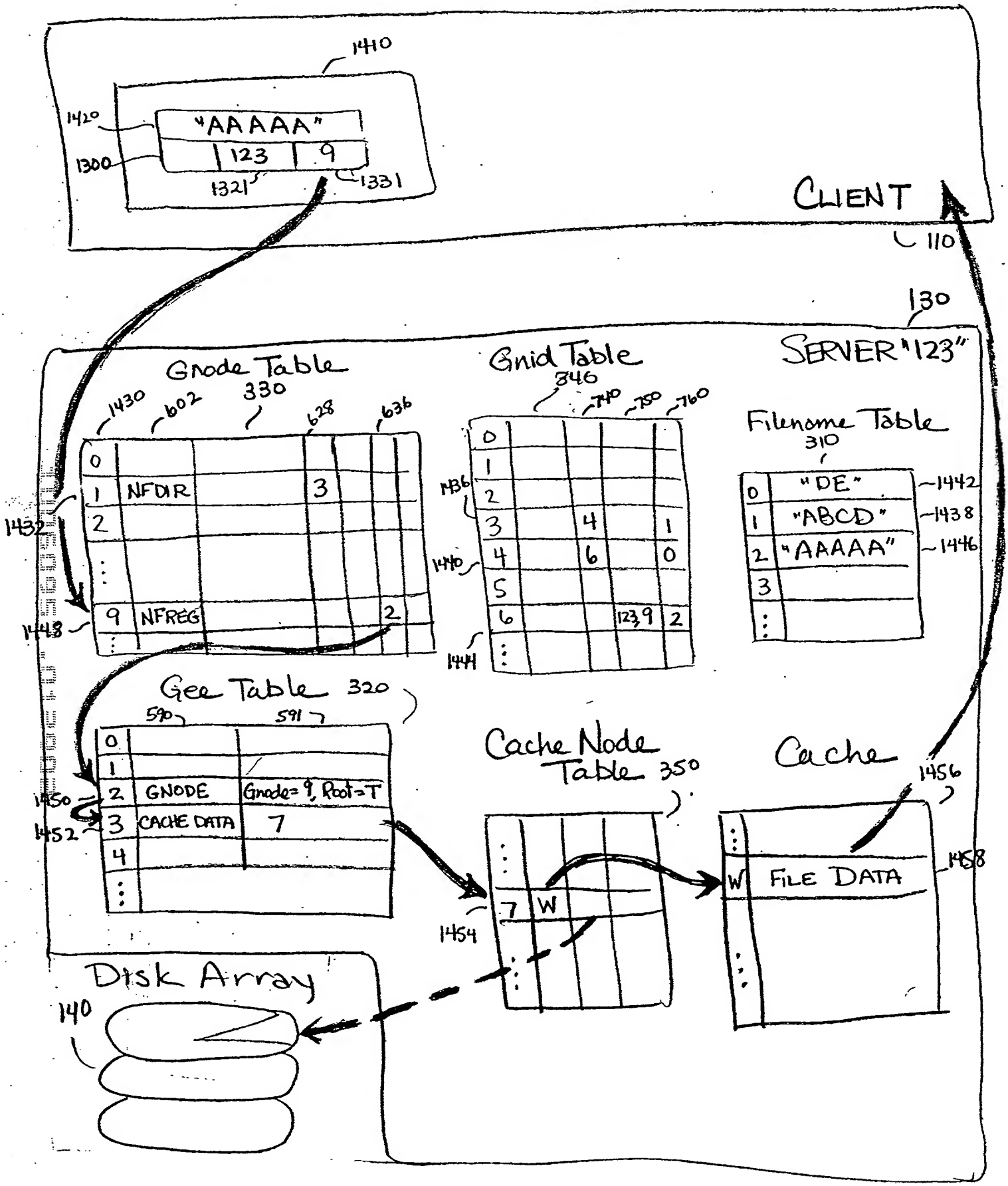


FIGURE 14b Example of a File Access

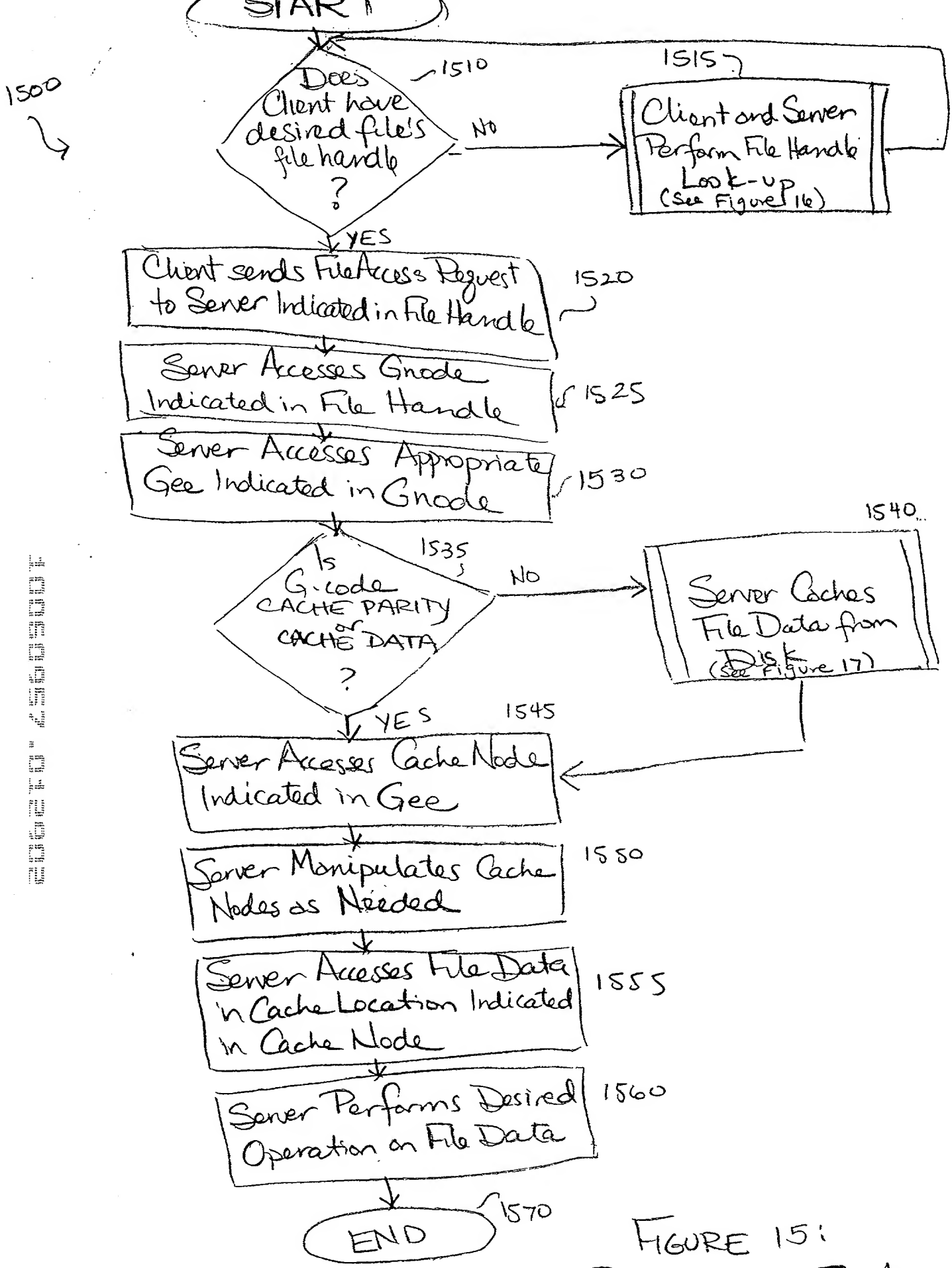


FIGURE 15:
Performing a File Access

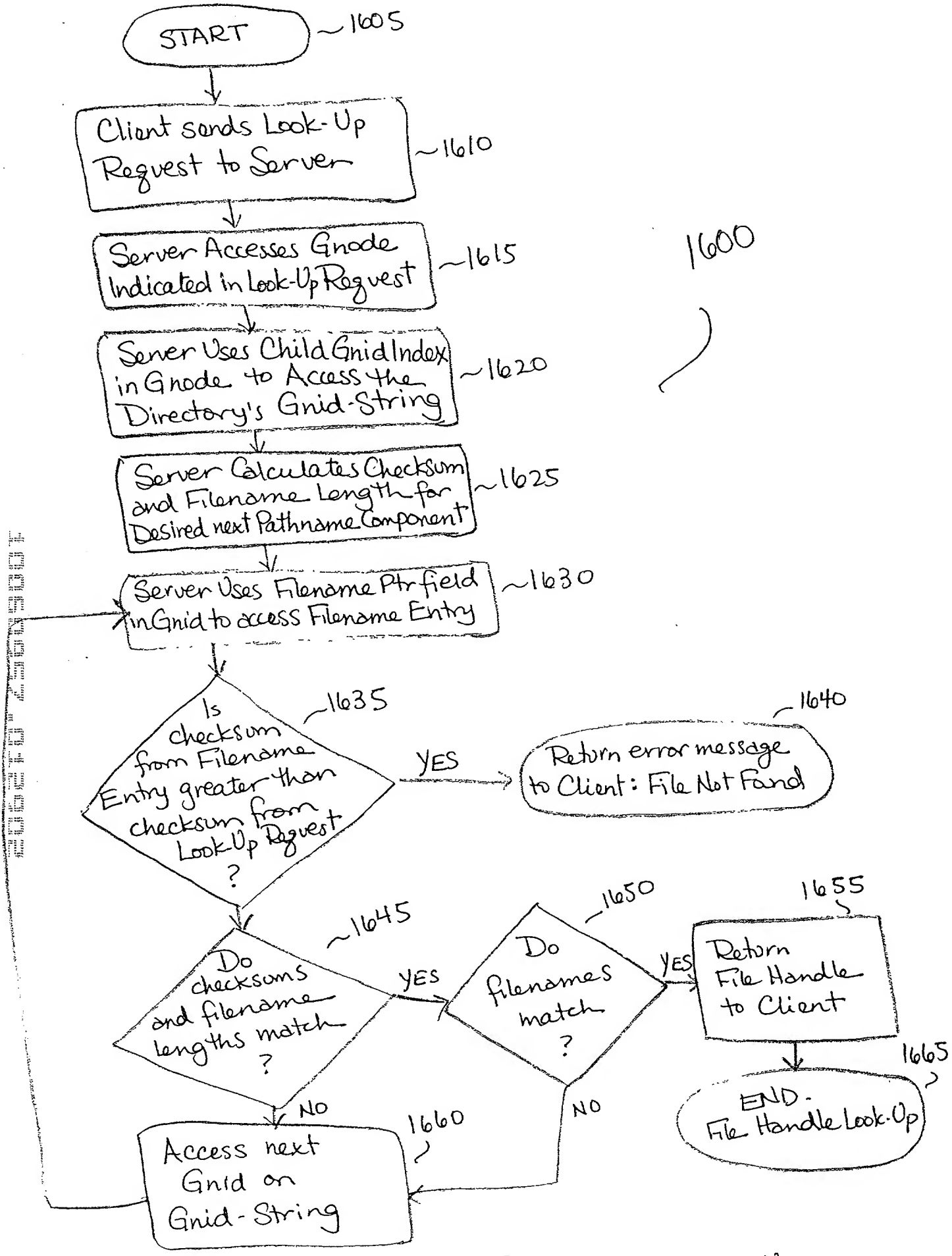


FIGURE 16: Performing a File Handle Look-Up

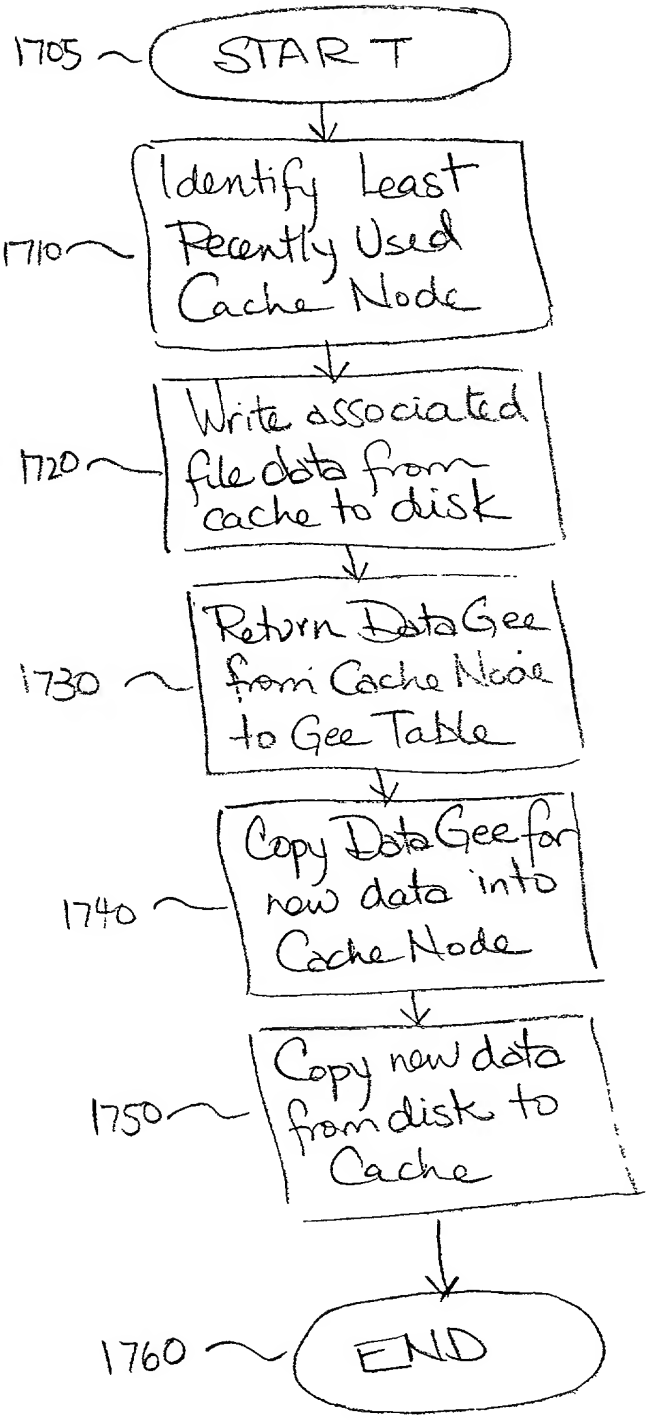


FIGURE 17: Caching File Data

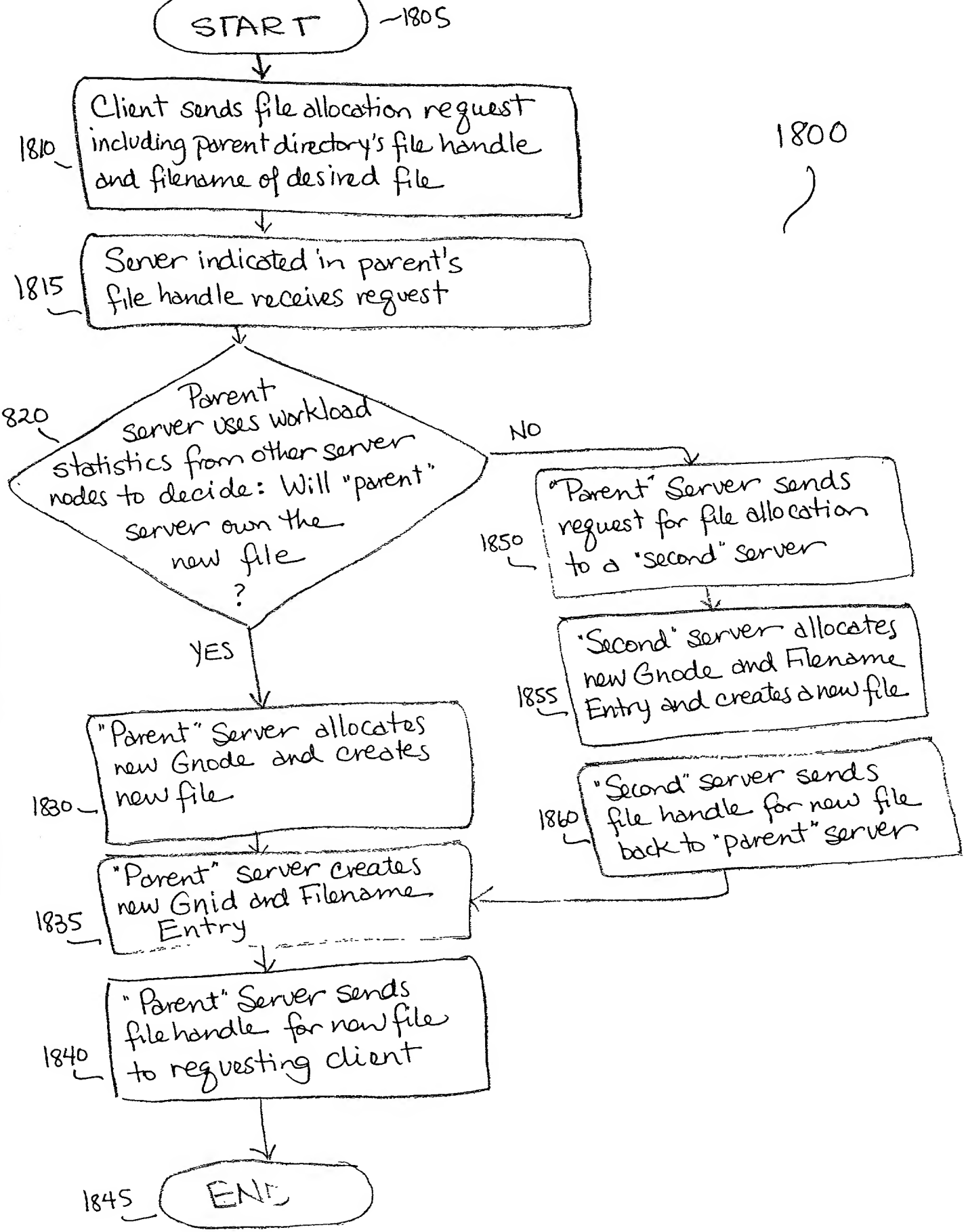


FIGURE 18 - File Allocation

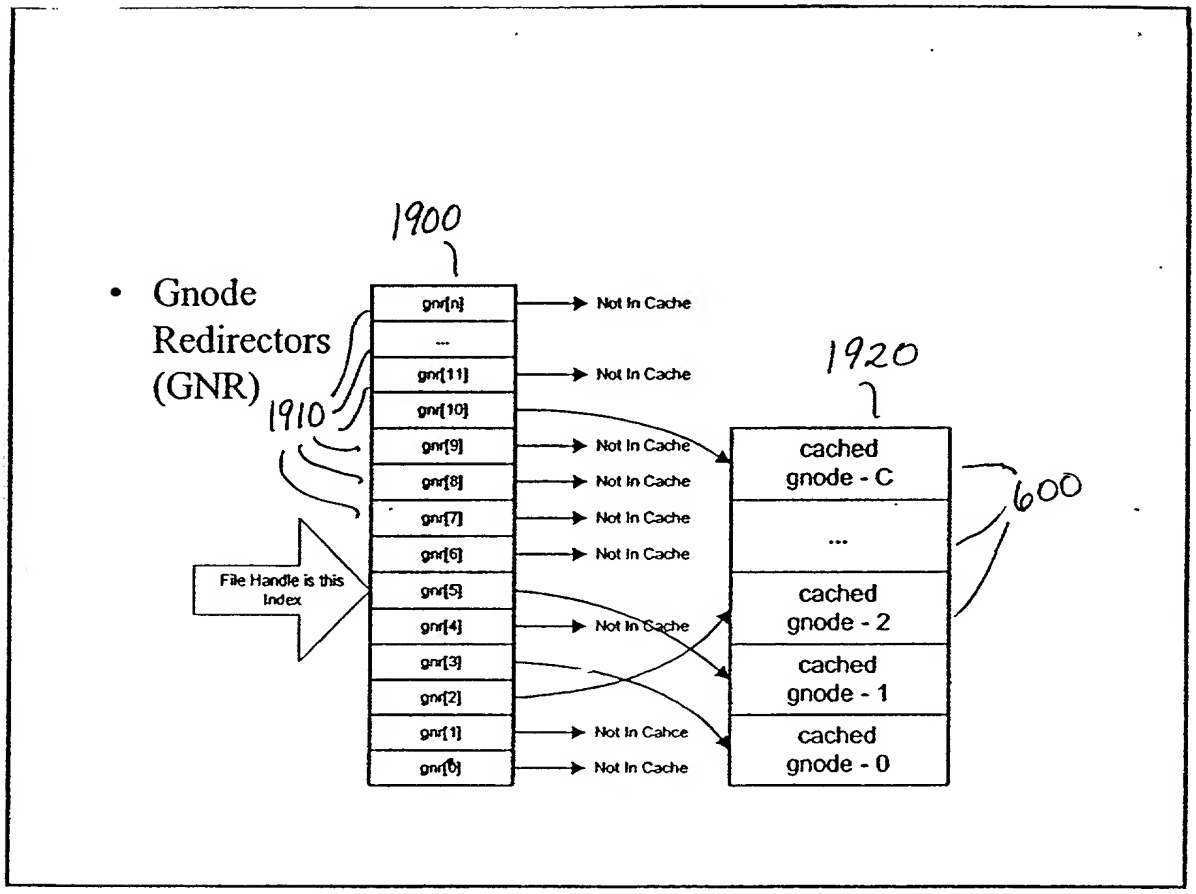

• Gnode Redirectors (GNR)
1900
1910
1920
600
gnr[n]
...
gnr[11]
gnr[10]
gnr[9]
gnr[8]
gnr[7]
gnr[6]
gnr[5]
gnr[4]
gnr[3]
gnr[2]
gnr[1]
gnr[0]
Not In Cache
Not In Cache
Not In Cache
Not In Cache
Not In Cache
Not In Cache
Not In Cahce
Not In Cache
File Handle is this Index
cached gnode - C
...
cached gnode - 2
cached gnode - 1
cached gnode - 0

FIGURE 19

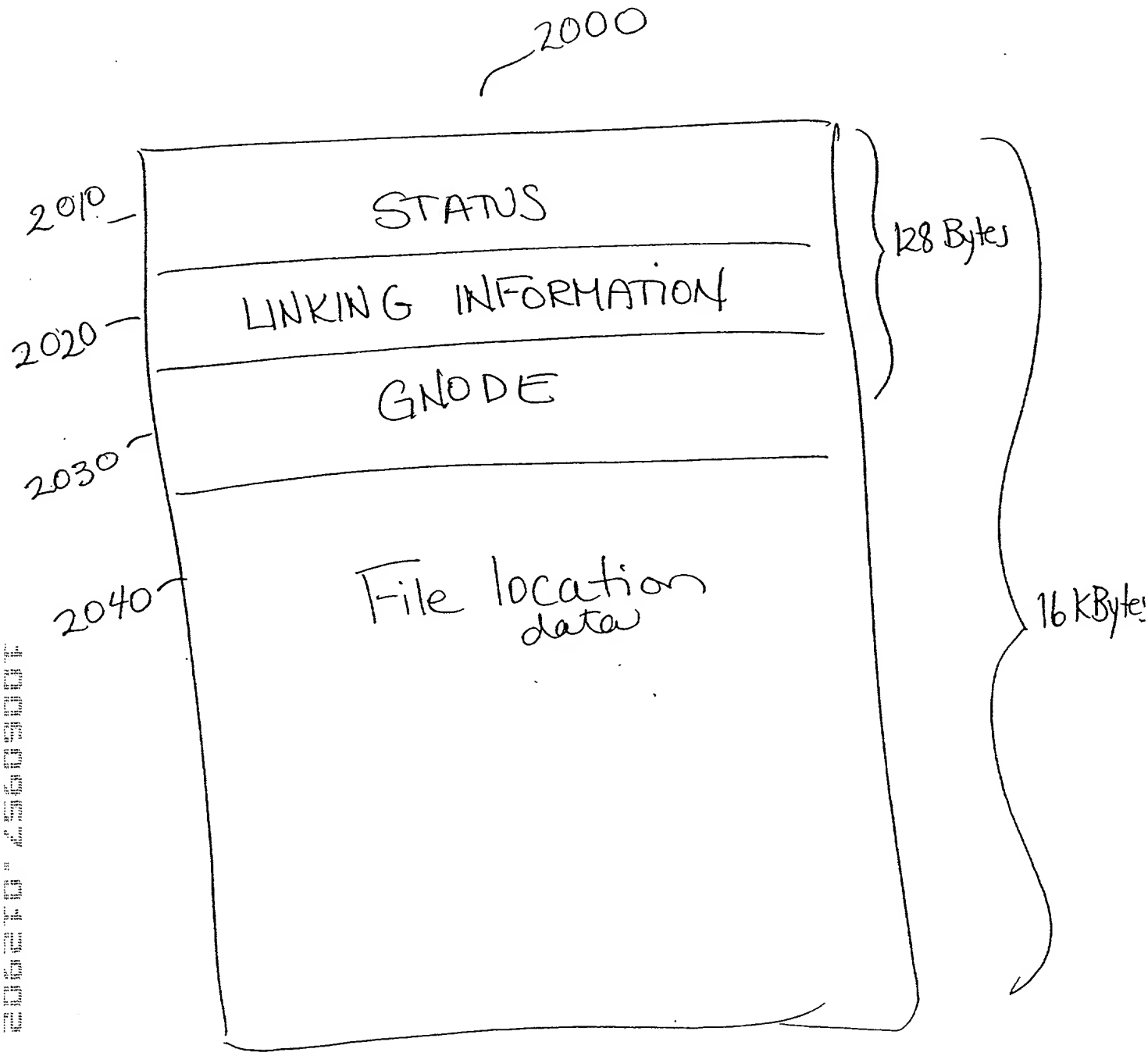

2000
2010
STATUS
2020
LINKING INFORMATION
2030
GNODE
2040
File location data
128 Bytes
16 KBytes

Figure 20a

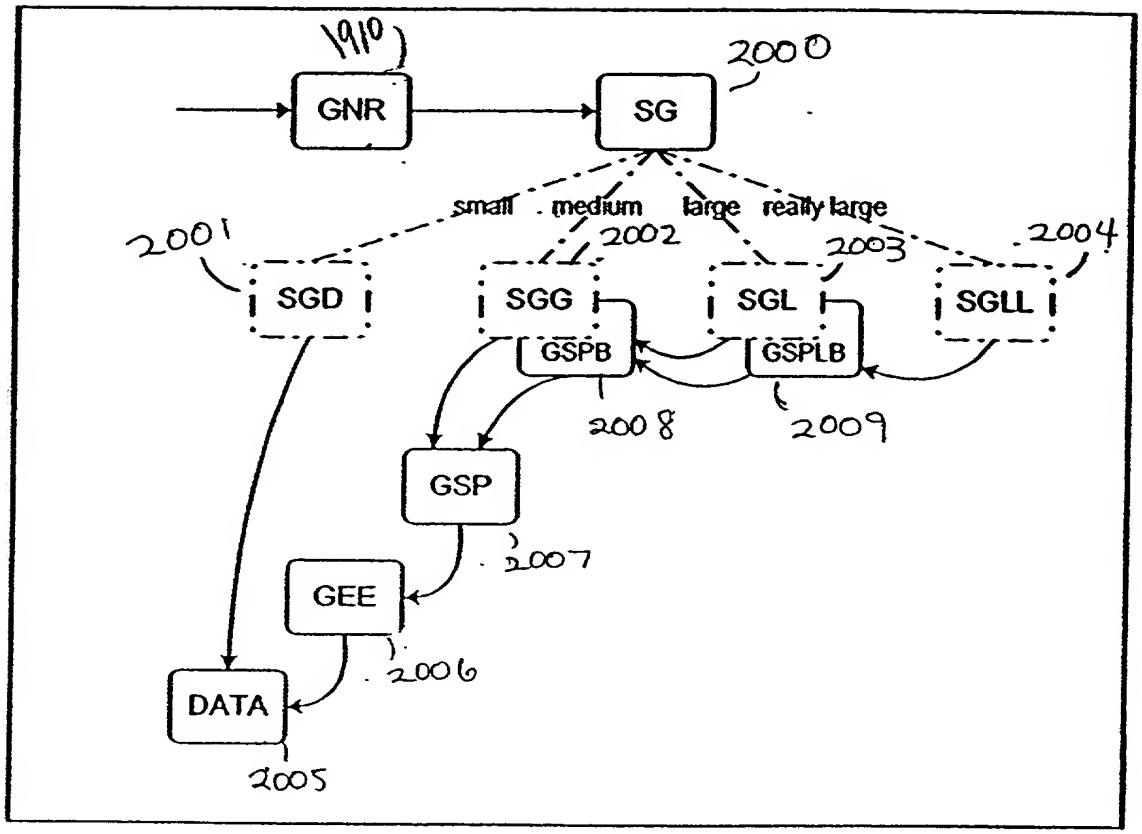

1910
GNR
2000
SG
small
medium
large
really large
2001
2002
2003
2004
SGD
SGG
SGL
SGLL
GSPB
GSPLB
2008
2009
GSP
2007
GEE
2006
DATA
2005

FIGURE 20b

CONVENTIONAL RAID MAPPING
(PRIOR ART)

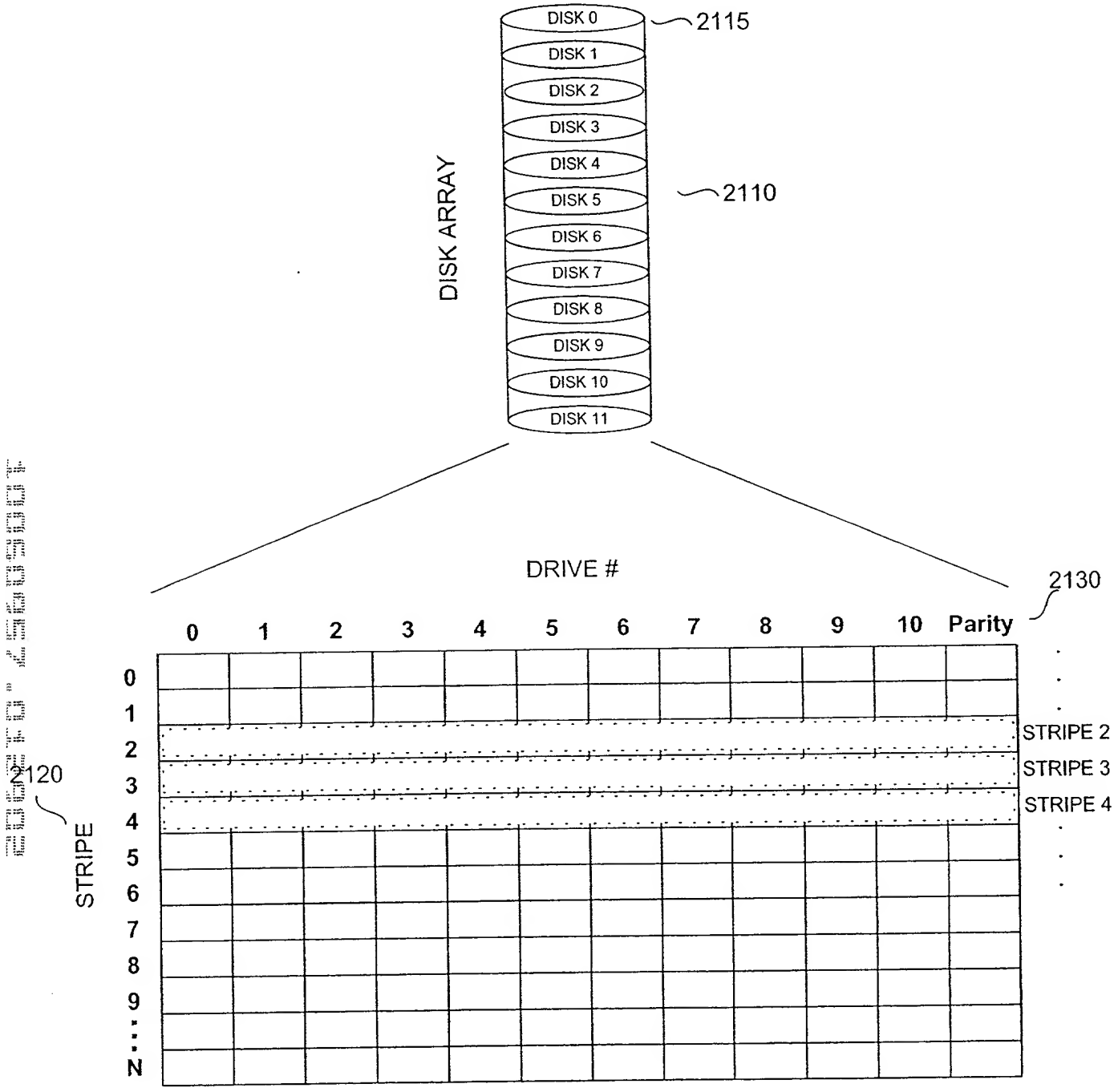


FIGURE 21

FIGURE 22A

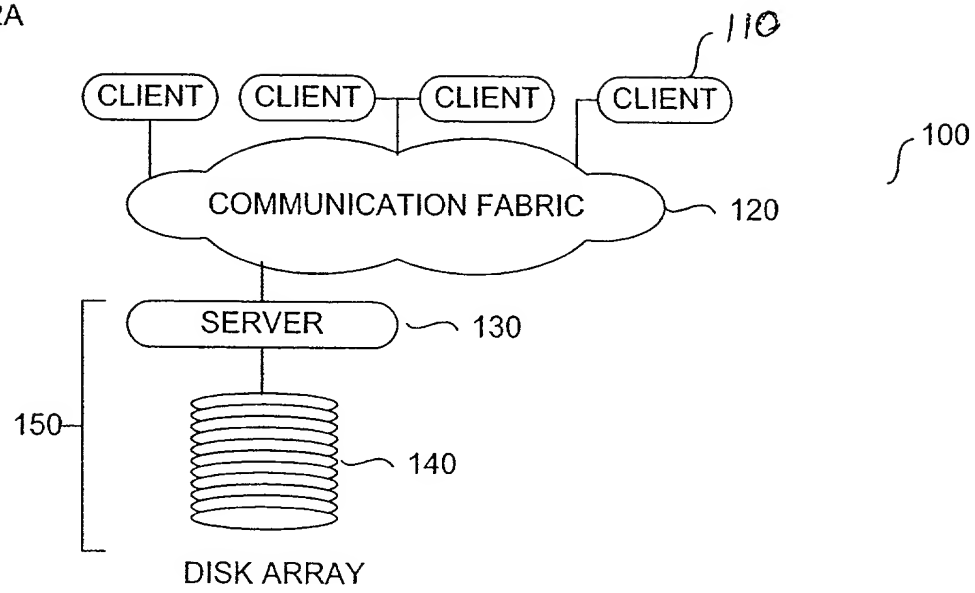


FIGURE 22B

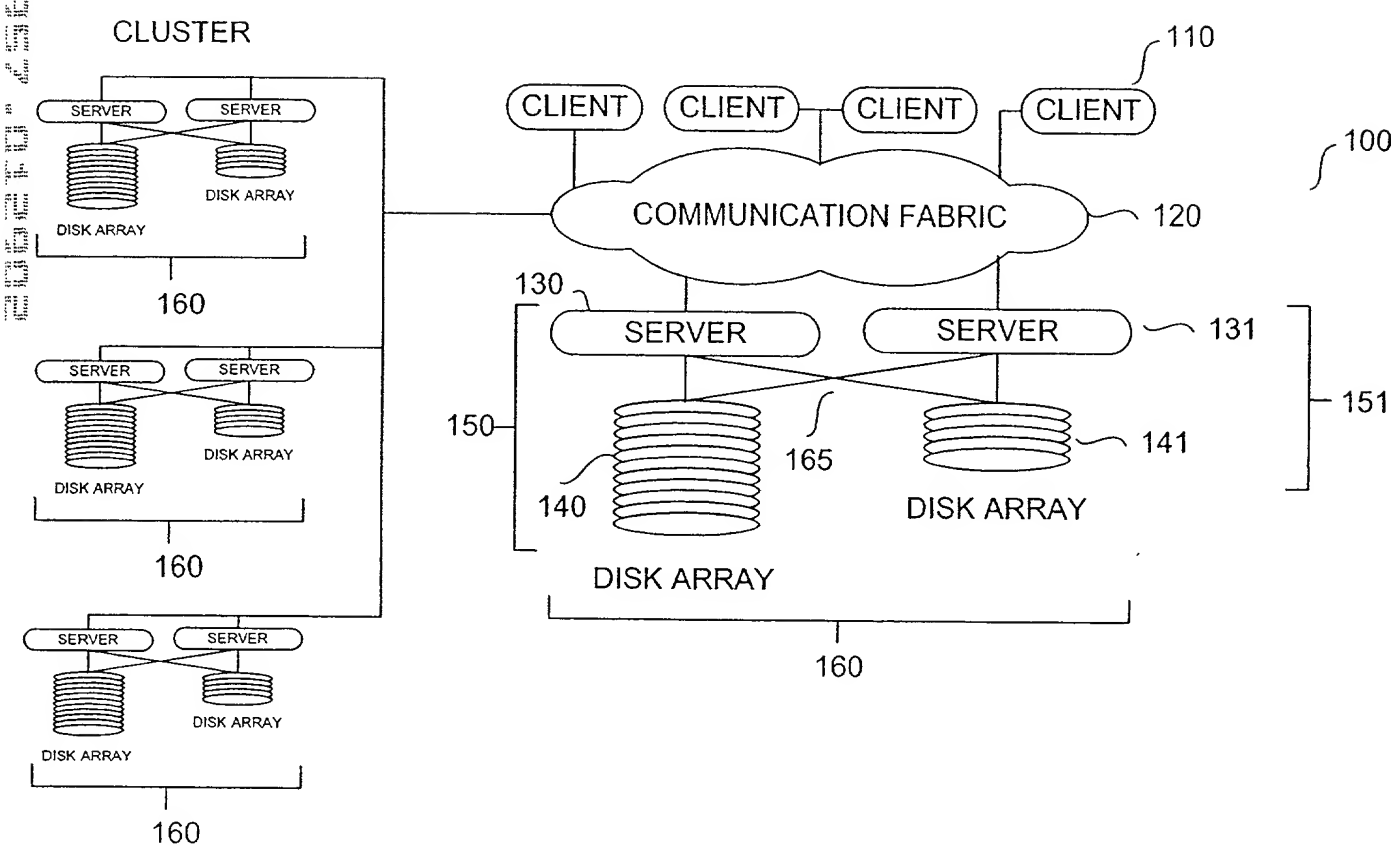


FIGURE 23

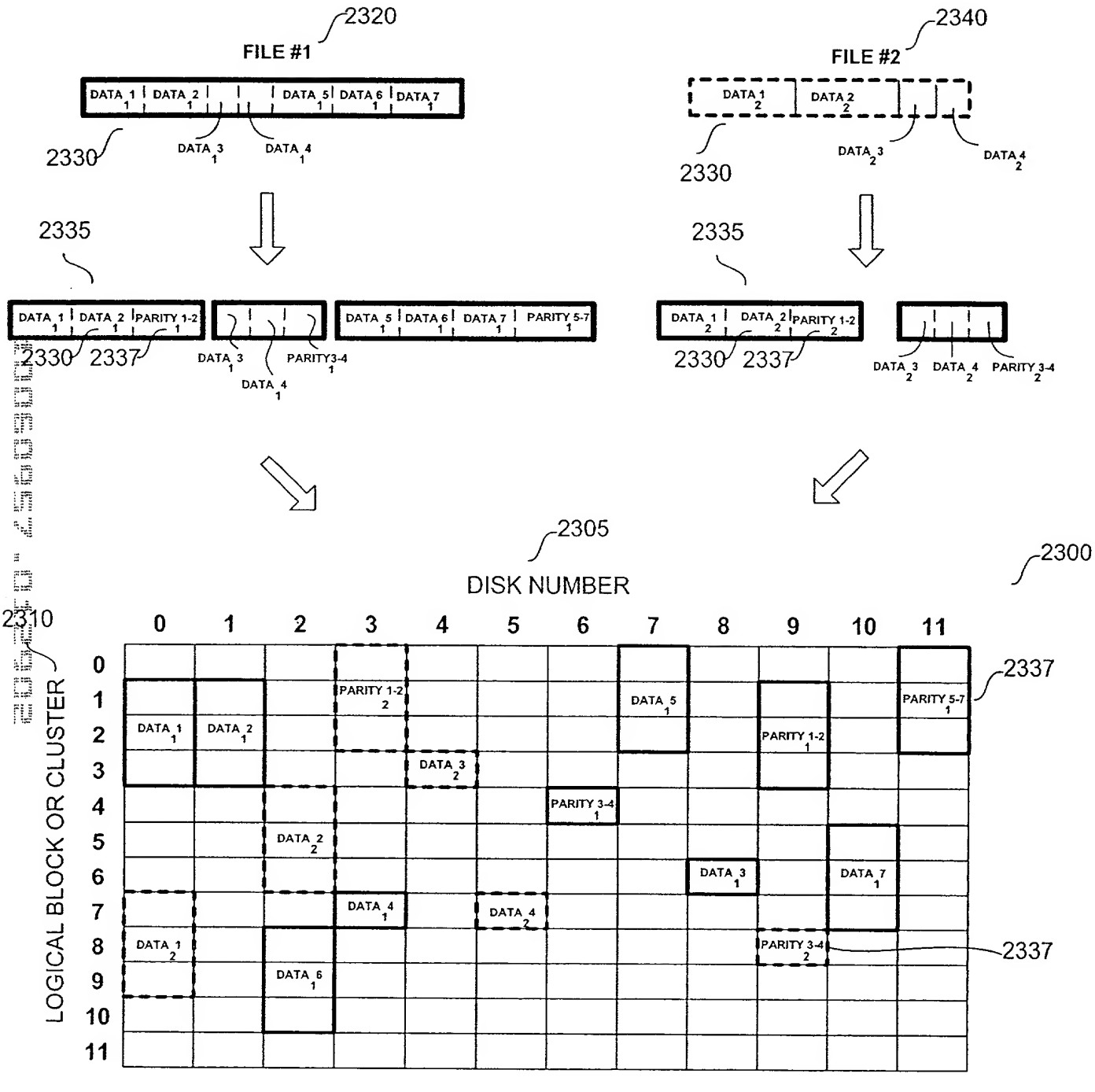

2320
FILE #1
DATA 1 1
DATA 2 1
DATA 5 1
DATA 6 1
DATA 7 1
2330
DATA 3 1
DATA 4 1
2340
FILE #2
DATA 1 2
DATA 2 2
DATA 3 2
DATA 4 2
2330
2335
DATA 1 1
DATA 2 1
PARITY 1-2 1
2330
2337
DATA 3 1
PARITY 3-4 1
DATA 4 1
DATA 5 1
DATA 6 1
DATA 7 1
PARITY 5-7 1
2335
DATA 1 2
DATA 2 2
PARITY 1-2 2
2330
2337
DATA 3 2
DATA 4 2
PARITY 3-4 2
2305
DISK NUMBER
2300
2310
LOGICAL BLOCK OR CLUSTER
0 1 2 3 4 5 6 7 8 9 10 11
PARITY 1-2 2
DATA 5 1
PARITY 5-7 1
2337
DATA 1 1
DATA 2 1
DATA 3 2
PARITY 1-2 1
PARITY 3-4 1
DATA 2 2
DATA 3 1
DATA 7 1
DATA 4 1
DATA 4 2
DATA 1 2
PARITY 3-4 2
2337
DATA 6 1

FIGURE 24A

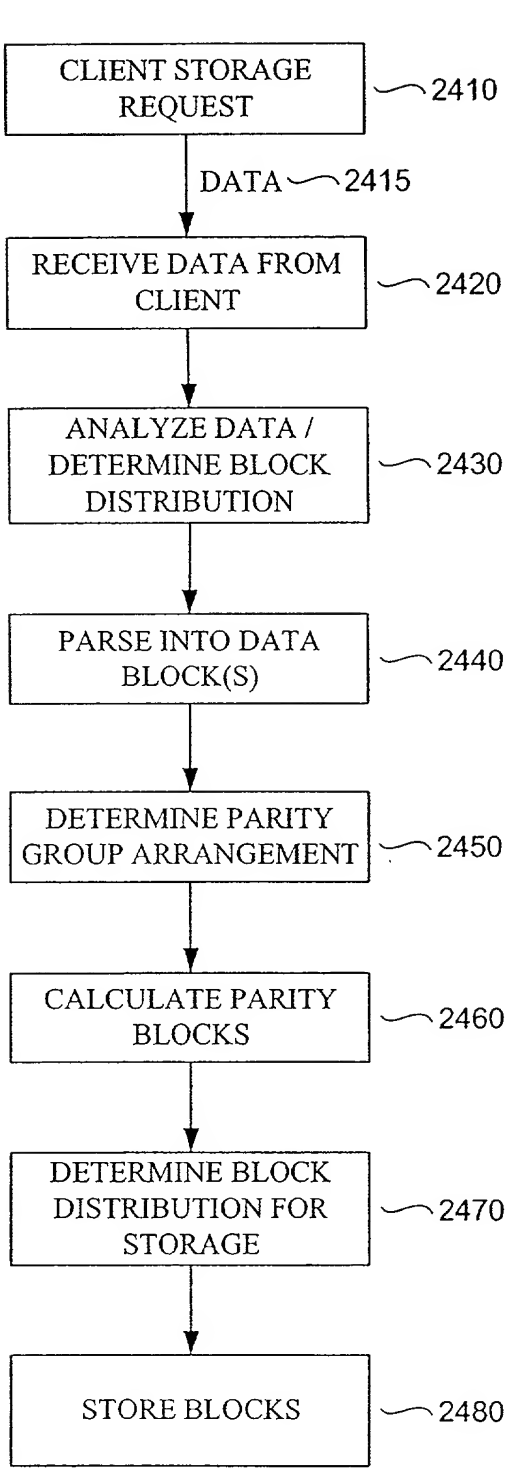

2400
CLIENT STORAGE REQUEST
2410
DATA
2415
RECEIVE DATA FROM CLIENT
2420
ANALYZE DATA / DETERMINE BLOCK DISTRIBUTION
2430
PARSE INTO DATA BLOCK(S)
2440
DETERMINE PARITY GROUP ARRANGEMENT
2450
CALCULATE PARITY BLOCKS
2460
DETERMINE BLOCK DISTRIBUTION FOR STORAGE
2470
STORE BLOCKS
2480

FIGURE 24B

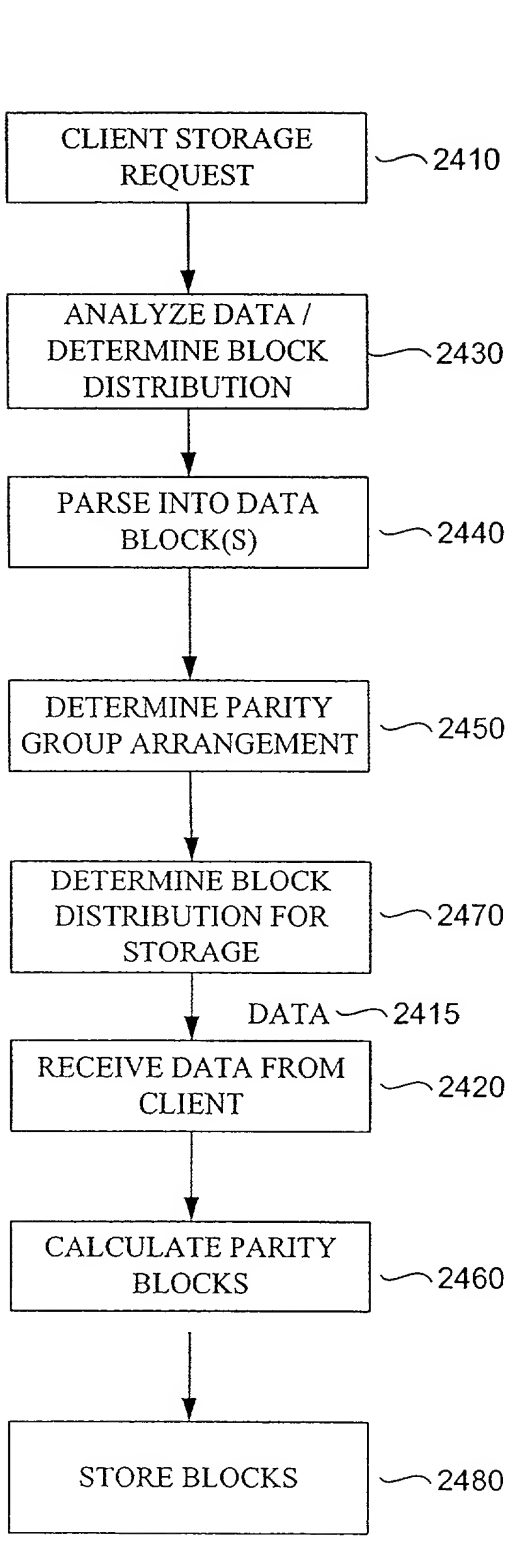

2405
CLIENT STORAGE REQUEST
2410
ANALYZE DATA / DETERMINE BLOCK DISTRIBUTION
2430
PARSE INTO DATA BLOCK(S)
2440
DETERMINE PARITY GROUP ARRANGEMENT
2450
DETERMINE BLOCK DISTRIBUTION FOR STORAGE
2470
DATA
2415
RECEIVE DATA FROM CLIENT
2420
CALCULATE PARITY BLOCKS
2460
STORE BLOCKS
2480

FIGURE 25

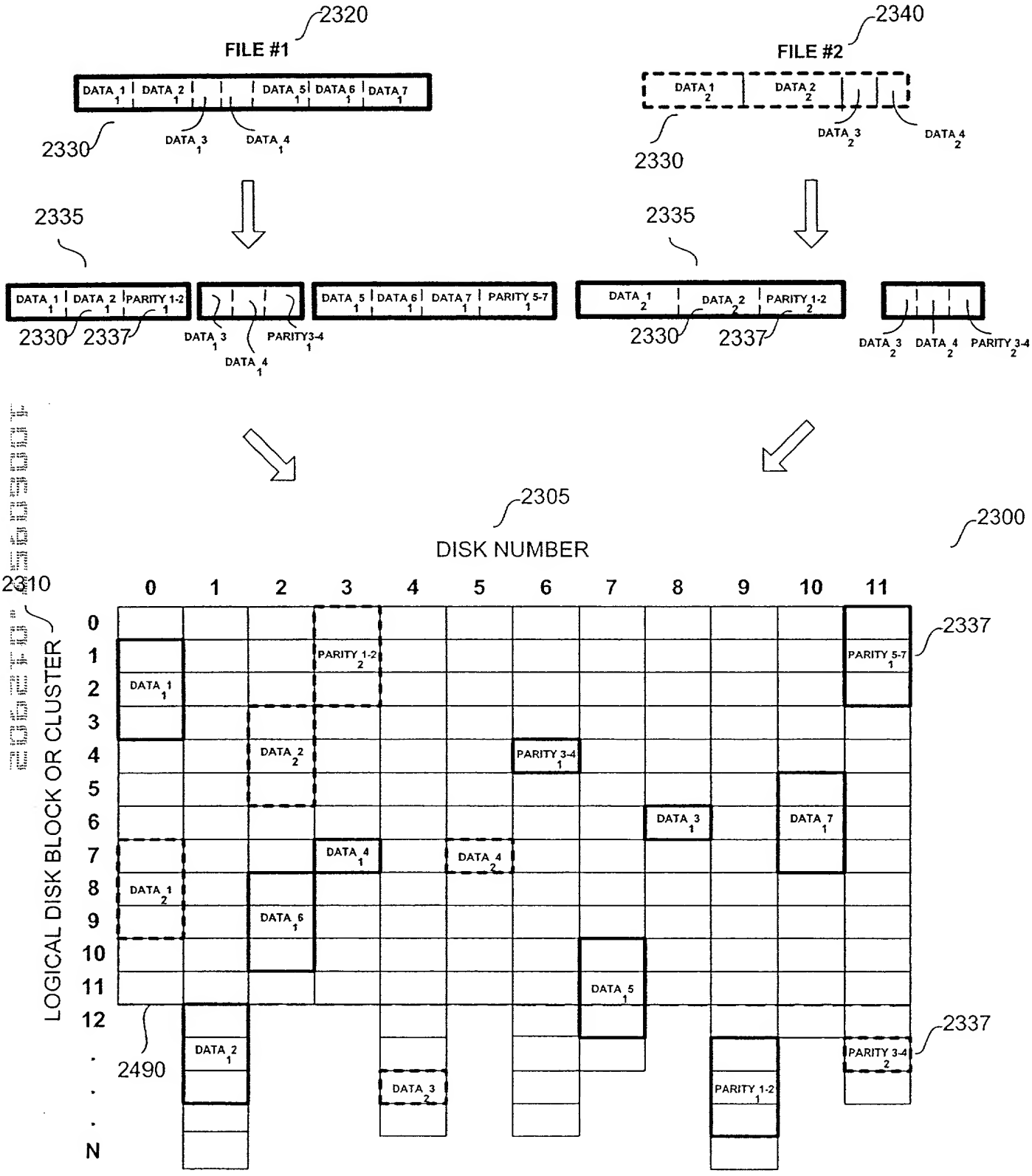

2320
FILE #1
DATA 1 1
DATA 2 1
DATA 5 1
DATA 6 1
DATA 7 1
2330
DATA 3 1
DATA 4 1
2340
FILE #2
DATA 1 2
DATA 2 2
DATA 3 2
DATA 4 2
2330
2335
DATA 1 1
DATA 2 1
PARITY 1-2 1
2330
2337
DATA 3 1
PARITY3-4 1
DATA 4 1
DATA 5 1
DATA 6 1
DATA 7 1
PARITY 5-7 1
2335
DATA 1 2
DATA 2 2
PARITY 1-2 2
2330
2337
DATA 3 2
DATA 4 2
PARITY 3-4 2
2305
DISK NUMBER
2300
2310
LOGICAL DISK BLOCK OR CLUSTER
0 1 2 3 4 5 6 7 8 9 10 11
0
1
2
3
4
5
6
7
8
9
10
11
12
.
.
.
N
PARITY 1-2 2
DATA 1 1
DATA 2 2
PARITY 3-4 1
DATA 3 1
DATA 7 1
DATA 4 1
DATA 4 2
DATA 1 2
DATA 6 1
DATA 5 1
DATA 2 1
DATA 3 2
PARITY 1-2 1
2337
PARITY 5-7 1
2337
PARITY 3-4 2
2490

FIGURE 26A

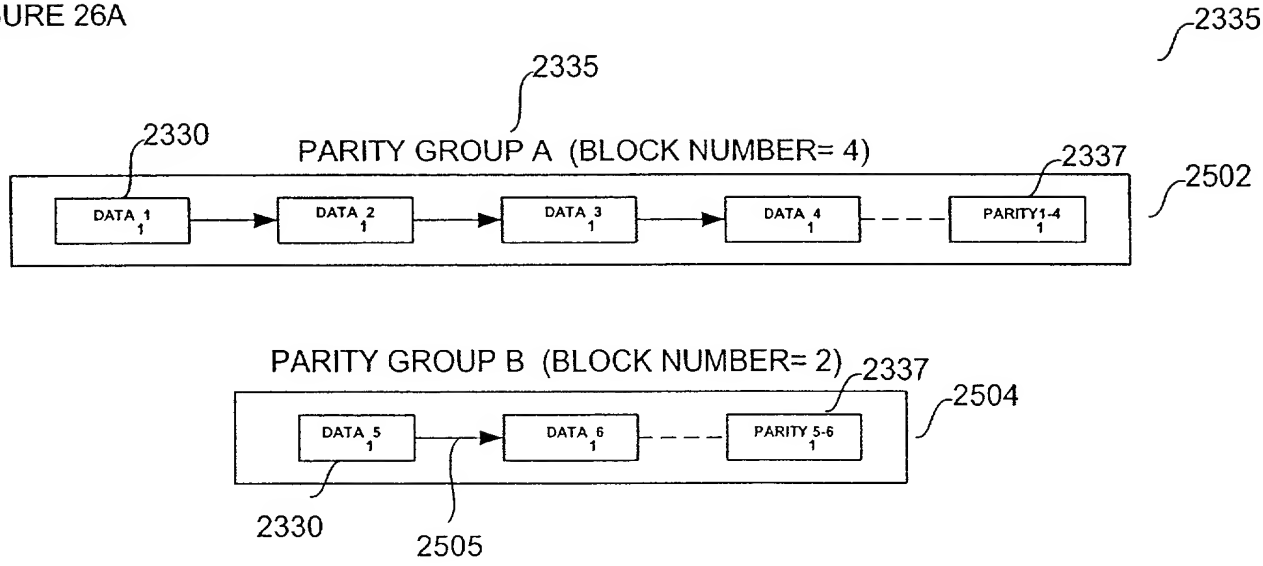
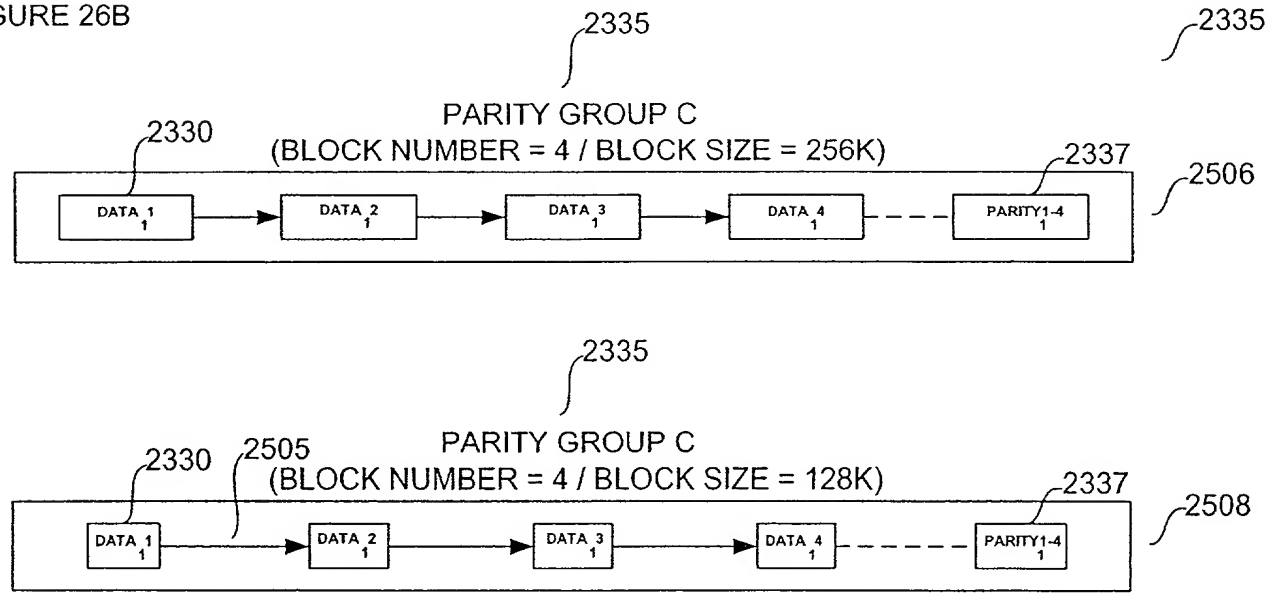


FIGURE 26B



DISK ARRAY INITIALIZATION USING GEE TABLE SPACE ALLOCATION

2530

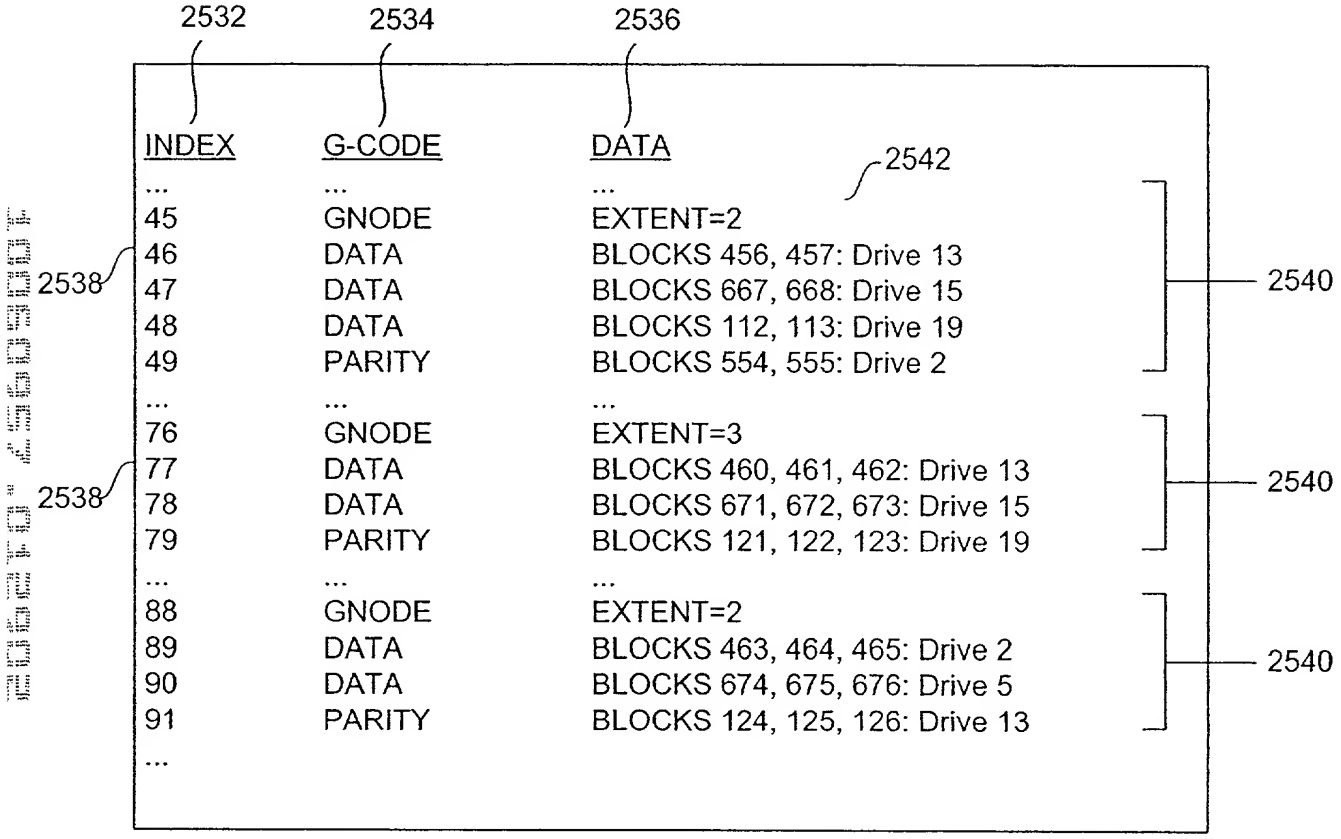


FIGURE 27

ARRAY PREPARATION / G-TABLE FORMATTING

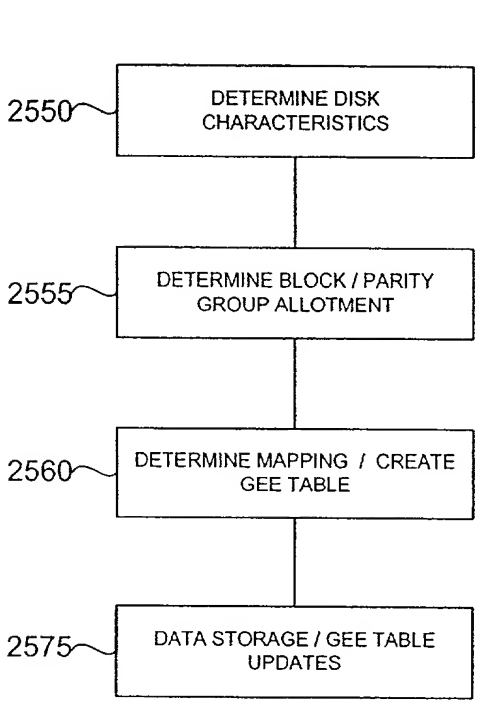
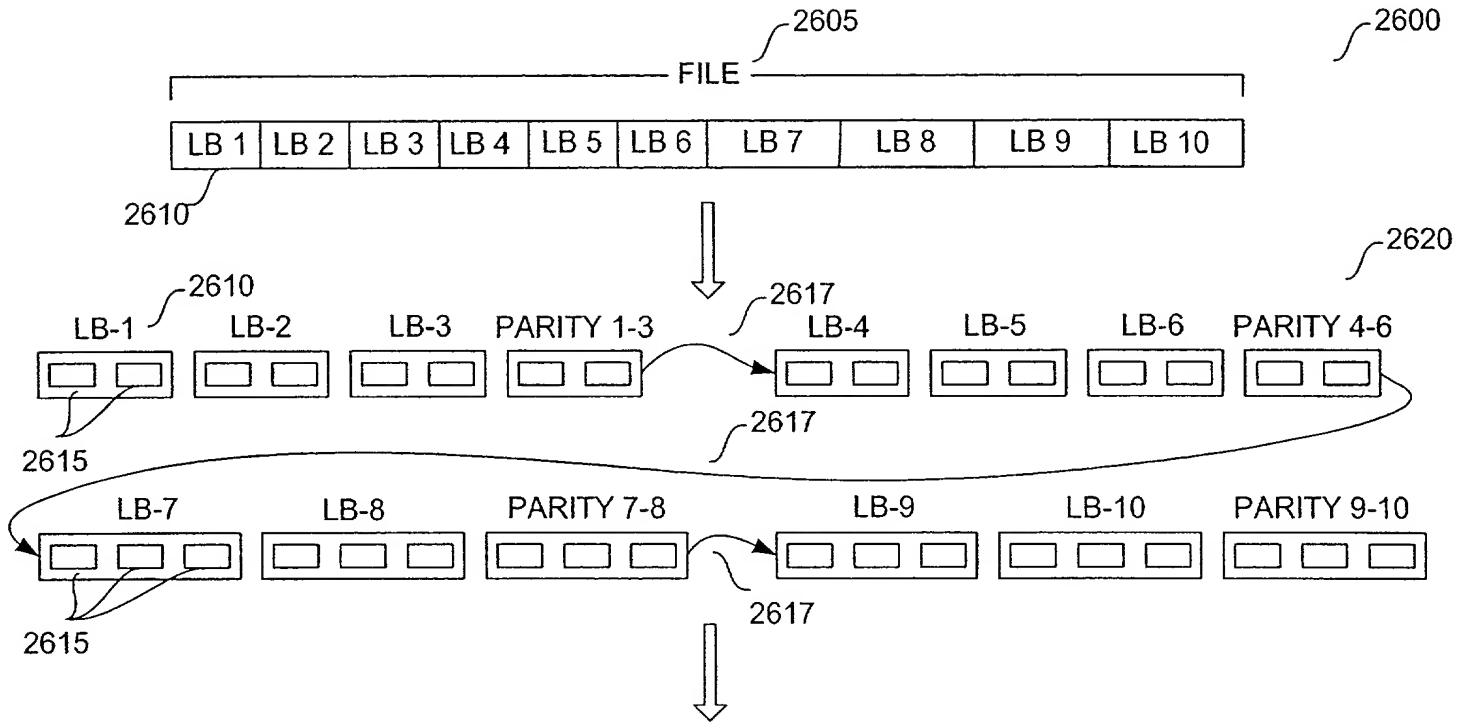


FIGURE 28



2530

2532 INDEX	2534 G-CODE	2536 DATA	FILE LOGICAL BLOCK
...	...	...	
45	GNODE	GNODE = 67, EXTENT=2, ROOT = TRUE	
46	DATA	BLOCKS 456, 457: Drive 13	1
47	DATA	BLOCKS 667, 668: Drive 15	2
48	DATA	BLOCKS 112, 113: Drive 19	3
49	PARITY	BLOCKS 554, 555: Drive 2	
50	DATA	BLOCKS 458, 459: Drive 13	4
51	DATA	BLOCKS 669, 670: Drive 15	5
52	DATA	BLOCKS 114, 115: Drive 19	6
53	PARITY	BLOCKS 556, 557: Drive 2	
54	LINK	INDEX 76	
...	...	...	
76	GNODE	GNODE = 67, EXTENT=3, ROOT = FALSE	
77	DATA	BLOCKS 460, 461, 462: Drive 13	7
78	DATA	BLOCKS 671, 672, 673: Drive 15	8
79	PARITY	BLOCKS 121, 122, 123: Drive 19	
80	LINK	INDEX 88	
...	...	...	
88	GNODE	GNODE = 67, EXTENT=2, ROOT = FALSE	
89	DATA	BLOCKS 463, 464, 465: Drive 2	9
90	DATA	BLOCKS 674, 675, 676: Drive 5	10
91	PARITY	BLOCKS 124, 125, 126: Drive 13	
...			

2635 2542 2637

2615

2655 2656 2657 2658

2627 2628 2629

2632 2632 2632

2630 2538 2660 2640 2661 2650

FIGURE 29

DRIVE FAILURE RECOVERY MECHANISM

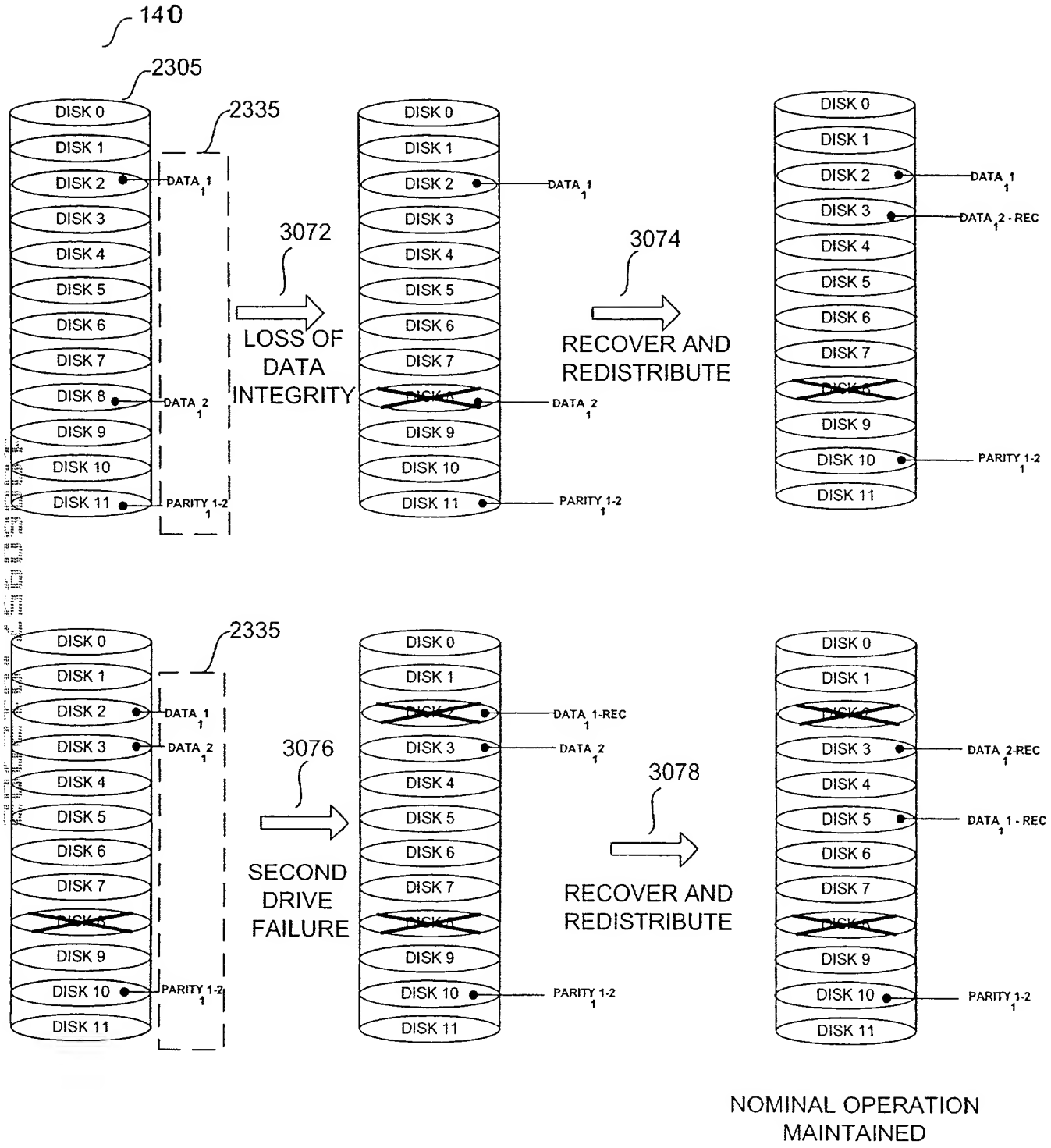
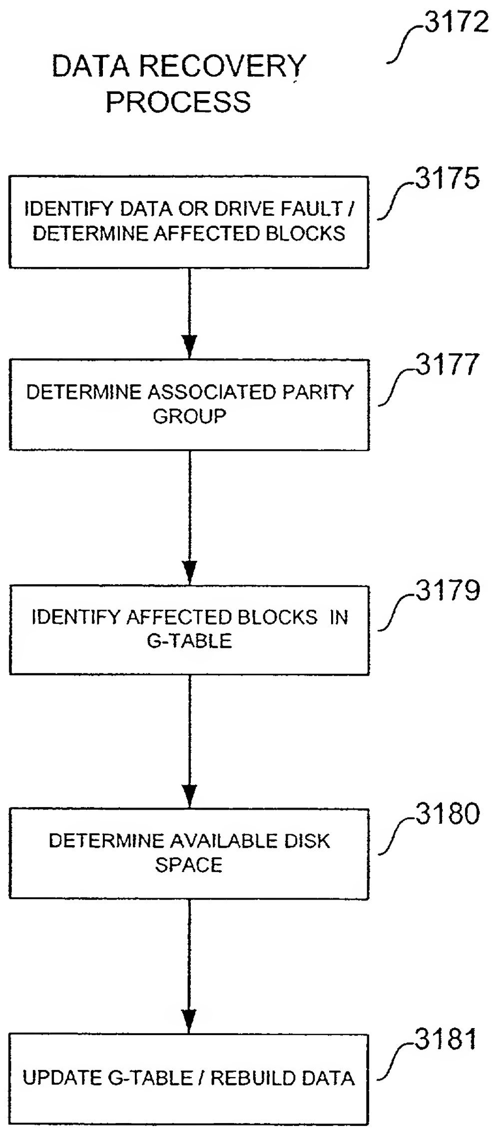


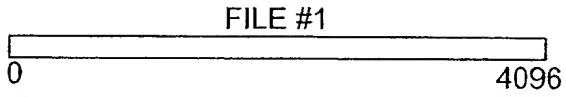
FIGURE 30



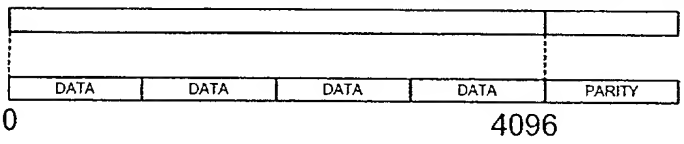
DATA RECOVERY PROCESS
3172
IDENTIFY DATA OR DRIVE FAULT / DETERMINE AFFECTED BLOCKS
3175
DETERMINE ASSOCIATED PARITY GROUP
3177
IDENTIFY AFFECTED BLOCKS IN G-TABLE
3179
DETERMINE AVAILABLE DISK SPACE
3180
UPDATE G-TABLE / REBUILD DATA
3181

FIGURE 31

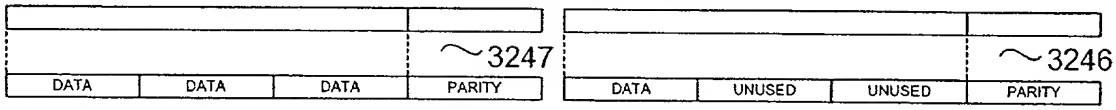
FIGURE 32A



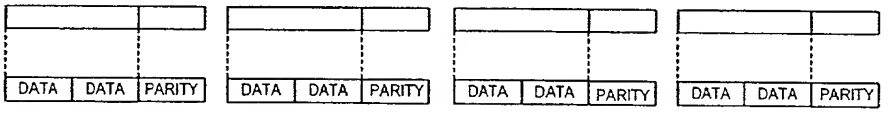
FILE #1 W/ PARITY -- 4-BLOCK PARITY GROUP -- EXTENT = 2 ~ 3240
5120 BYTES TOTAL / UTILIZATION = 100%



FILE #1 W/ PARITY -- 3-BLOCK PARITY GROUP -- EXTENT = 2 ~ 3241
8192 BYTES TOTAL / UTILIZATION = 66%



FILE #1 W/ PARITY -- 2-BLOCK PARITY GROUP -- EXTENT = 1 ~ 3242
6144 BYTES TOTAL / UTILIZATION = 100%



FILE #1 W/ PARITY -- 1-BLOCK PARITY GROUP -- EXTENT = 1 ~ 3243
8192 BYTES TOTAL / UTILIZATION = 100%

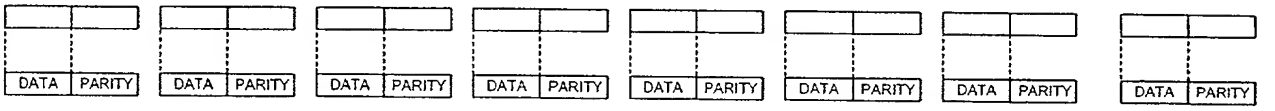
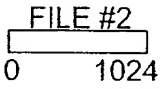
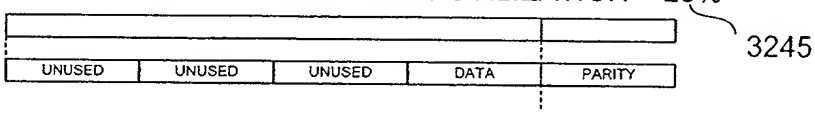


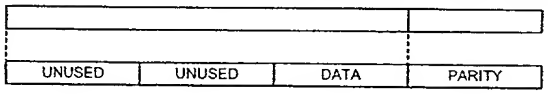
FIGURE 32B



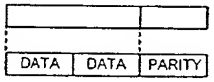
FILE #2 W/ PARITY -- 4-BLOCK PARITY GROUP -- EXTENT = 2 ~ 3250
5120 BYTES TOTAL / UTILIZATION = 25%



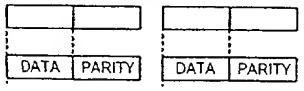
FILE #2 W/ PARITY -- 3-BLOCK PARITY GROUP -- EXTENT = 2 ~ 3251
4096 BYTES TOTAL / UTILIZATION = 33%



FILE #2 W/ PARITY -- 2-BLOCK PARITY GROUP -- EXTENT = 1 ~ 3252
1536 BYTES TOTAL / UTILIZATION = 100%



FILE #2 W/ PARITY -- 1-BLOCK PARITY GROUP -- EXTENT = 1 ~ 3253
2048 BYTES TOTAL / UTILIZATION = 100%



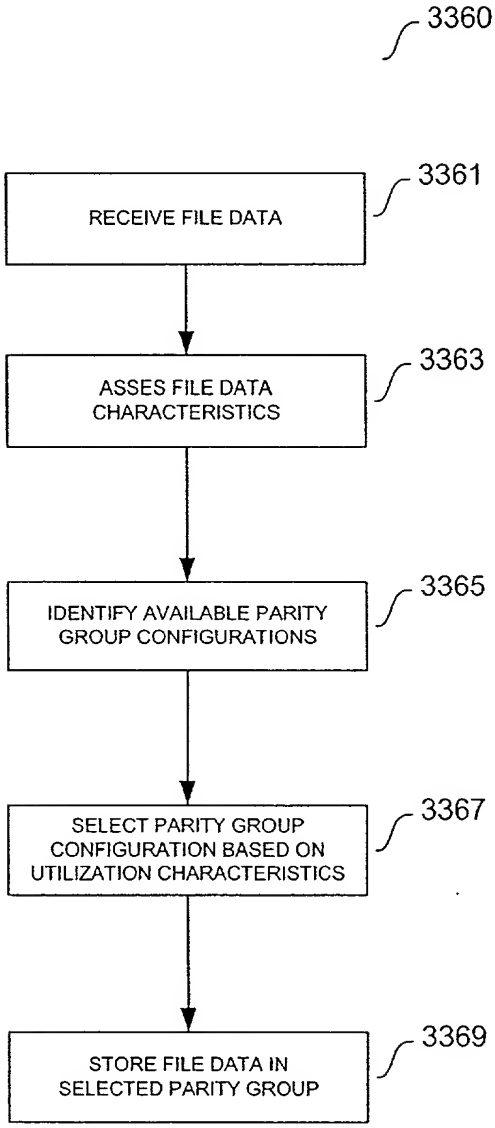

3360
RECEIVE FILE DATA
3361
ASSES FILE DATA CHARACTERISTICS
3363
IDENTIFY AVAILABLE PARITY GROUP CONFIGURATIONS
3365
SELECT PARITY GROUP CONFIGURATION BASED ON UTILIZATION CHARACTERISTICS
3367
STORE FILE DATA IN SELECTED PARITY GROUP
3369

FIGURE 33

FIGURE 34A

		INITIAL ALLOCATION (3491)	DISK SPACE % (3485)
DATA DATA DATA DATA PARITY	4 block parity (3480)	10000 groups	36%
DATA DATA DATA PARITY	3 block parity (3481)	10000 groups	28%
DATA DATA PARITY	2 block parity (3482)	10000 groups	22%
DATA PARITY	1 block parity (3483)	10000 groups	14%

DISK USAGE (3487)

FIGURE 34B

	FREE (3492)	OCCUPIED	TOTAL (3490)	DISK SPACE %
3480 4 block parity	2500 groups	7500 groups	10000 groups	36%
3481 3 block parity	7500 groups	2500 groups	10000 groups	28%
3482 2 block parity	3500 groups	6500 groups	10000 groups	22%
3483 1 block parity	500 groups	9500 groups	10000 groups	14%

REDISTRIBUTION (3494)

FIGURE 34C

		FREE (3492)	OCCUPIED	TOTAL (3490)	DISK SPACE %
3480 4 block parity		2500 groups	7500 groups	10000 groups	36%
3481 3 block parity	-5000 groups of 3 block parity	2500 groups	2500 groups	5000 groups	14%
3482 2 block parity		3500 groups	6500 groups	10000 groups	22%
3483 1 block parity	+10000 groups of 1 block parity	10500 groups	9500 groups	20000 groups	28% REDISTRIBUTION

PARITY GROUP REDISTRIBUTION PROCESSES 3500

3510

FIGURE 35A PARITY GROUP DISSOLUTION

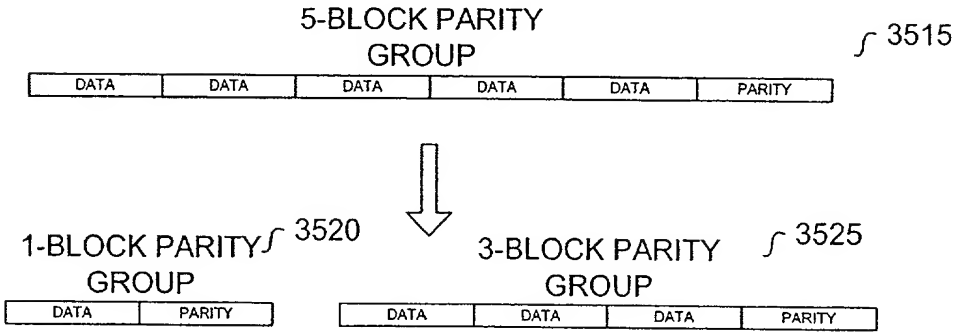
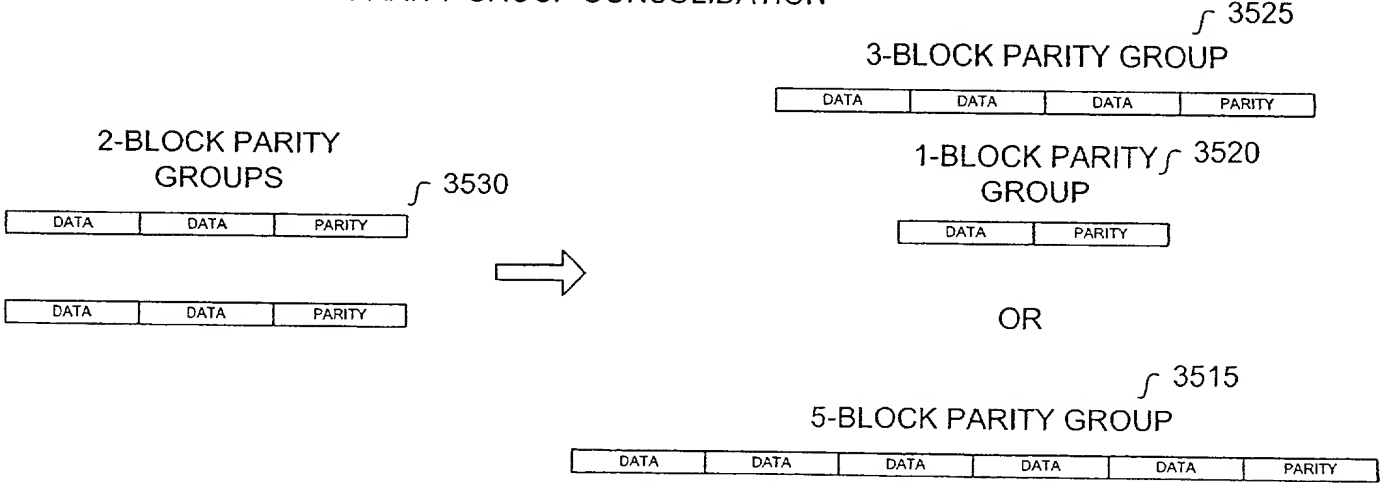


FIGURE 35B

3535

PARITY GROUP CONSOLIDATION



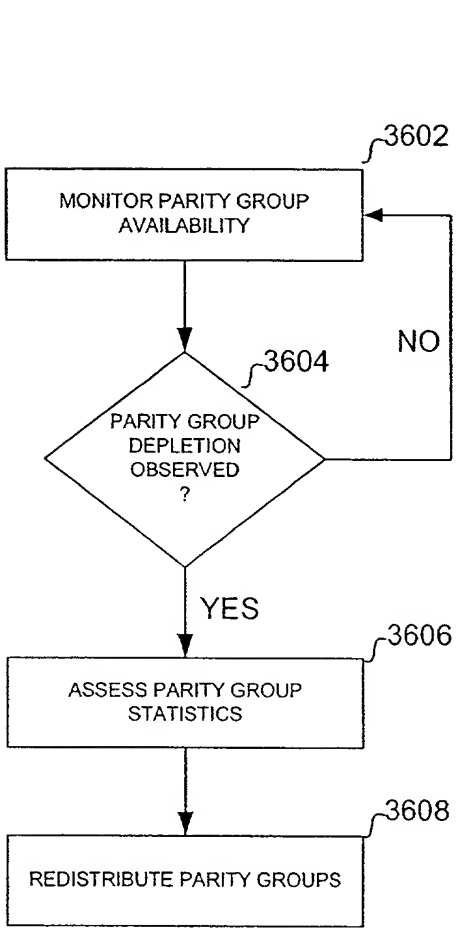

3600
3602
MONITOR PARITY GROUP AVAILABILITY
NO
3604
PARITY GROUP DEPLETION OBSERVED ?
YES
3606
ASSESS PARITY GROUP STATISTICS
3608
REDISTRIBUTE PARITY GROUPS

FIGURE 36

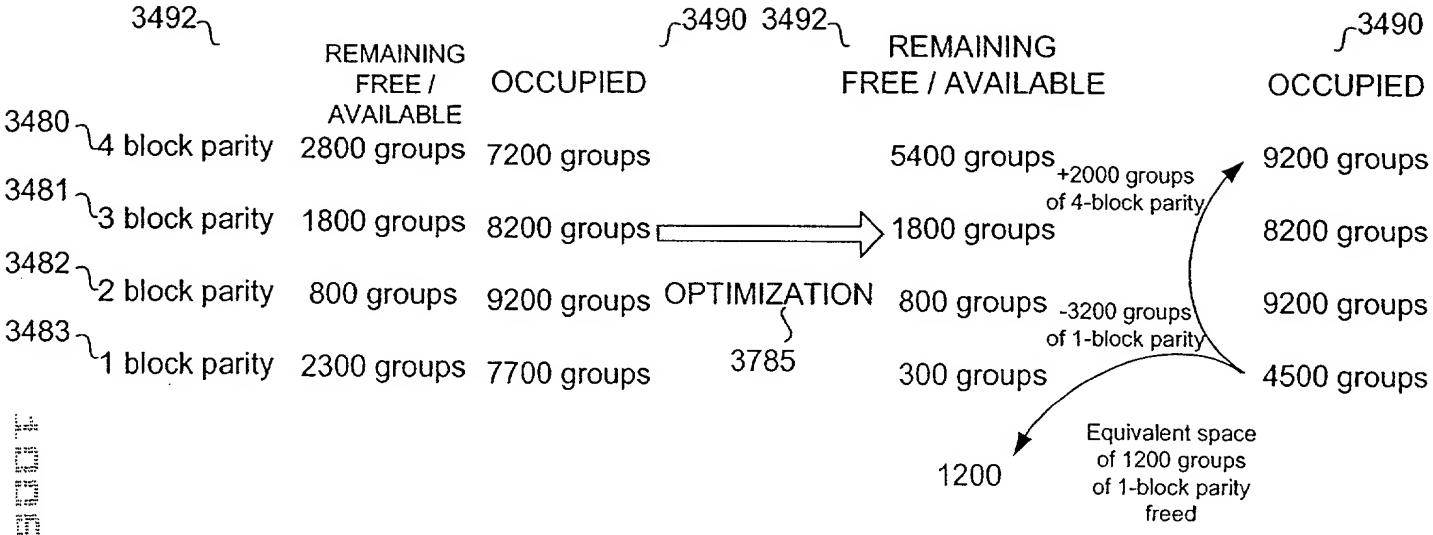

3492
REMAINING FREE / AVAILABLE
OCCUPIED
3490
3492
REMAINING FREE / AVAILABLE
3490
OCCUPIED
3480
4 block parity
2800 groups
7200 groups
5400 groups
+2000 groups of 4-block parity
9200 groups
3481
3 block parity
1800 groups
8200 groups
1800 groups
8200 groups
3482
2 block parity
800 groups
9200 groups
OPTIMIZATION
800 groups
-3200 groups of 1-block parity
9200 groups
3483
1 block parity
2300 groups
7700 groups
3785
300 groups
4500 groups
1200
Equivalent space of 1200 groups of 1-block parity freed

FIGURE 37

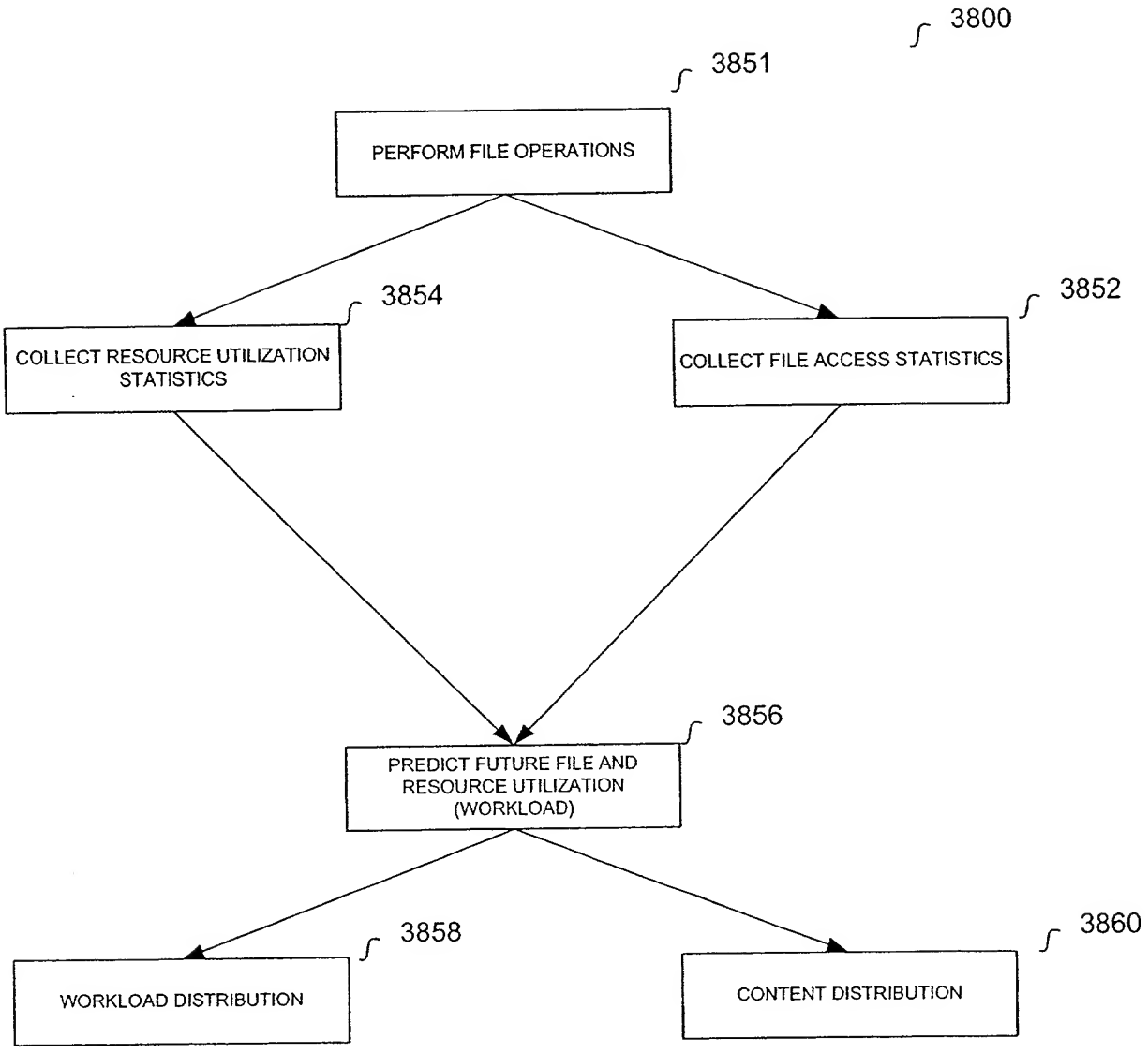

3800
3851
PERFORM FILE OPERATIONS
3854
COLLECT RESOURCE UTILIZATION STATISTICS
3852
COLLECT FILE ACCESS STATISTICS
3856
PREDICT FUTURE FILE AND RESOURCE UTILIZATION (WORKLOAD)
3858
WORKLOAD DISTRIBUTION
3860
CONTENT DISTRIBUTION

FIGURE 38

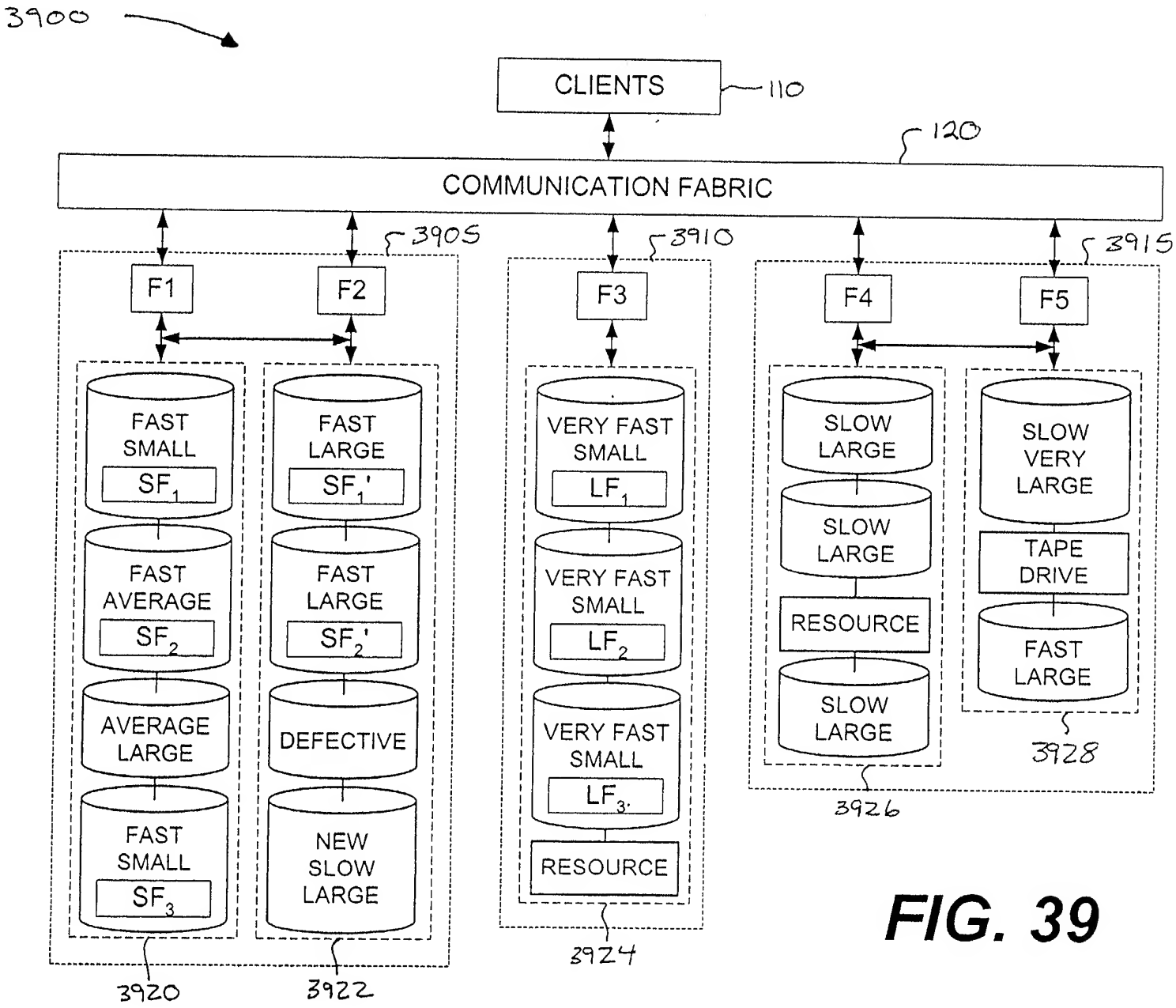

3900
CLIENTS
110
120
COMMUNICATION FABRIC
3905
3910
3915
F1
F2
F3
F4
F5
FAST SMALL
SF1
FAST AVERAGE
SF2
AVERAGE LARGE
FAST SMALL
SF3
FAST LARGE
SF1'
FAST LARGE
SF2'
DEFECTIVE
NEW SLOW LARGE
VERY FAST SMALL
LF1
VERY FAST SMALL
LF2
VERY FAST SMALL
LF3
RESOURCE
SLOW LARGE
SLOW LARGE
RESOURCE
SLOW LARGE
SLOW VERY LARGE
TAPE DRIVE
FAST LARGE
3920
3922
3924
3926
3928

FIG. 39

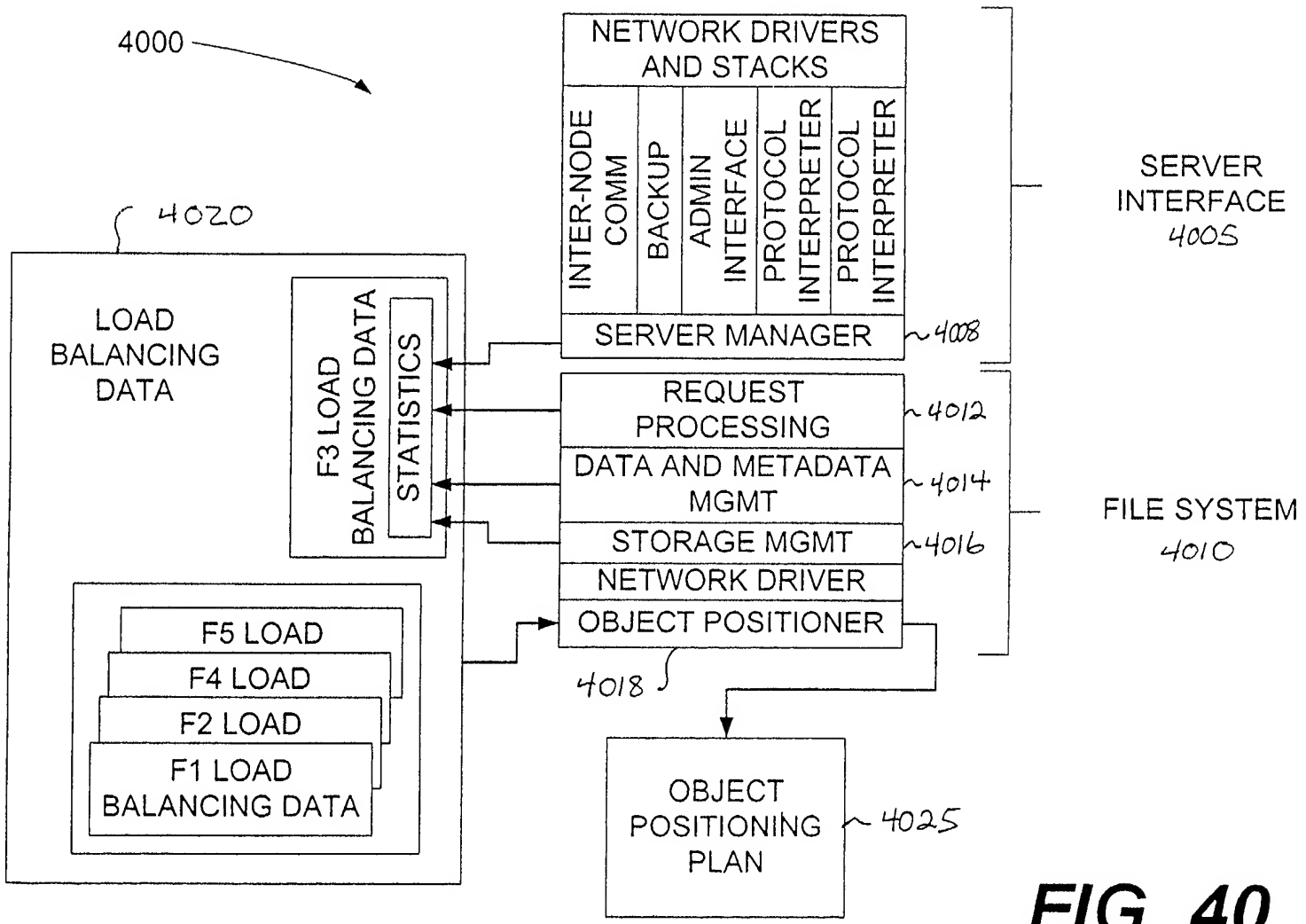

4000
NETWORK DRIVERS AND STACKS
INTER-NODE COMM
BACKUP
ADMIN INTERFACE
PROTOCOL INTERPRETER
PROTOCOL INTERPRETER
SERVER MANAGER
4008
SERVER INTERFACE
4005
4020
LOAD BALANCING DATA
F3 LOAD BALANCING DATA
STATISTICS
F5 LOAD
F4 LOAD
F2 LOAD
F1 LOAD BALANCING DATA
REQUEST PROCESSING
4012
DATA AND METADATA MGMT
4014
STORAGE MGMT
4016
NETWORK DRIVER
OBJECT POSITIONER
4018
FILE SYSTEM
4010
OBJECT POSITIONING PLAN
4025

FIG. 40

F3 OBJECT
POSITIONING PLAN

-Push LF to F4-F5 Cluster

-Issue File Handle For LF = Stale

-If Requested,
- -Send acceptance for copy of SF to F1
- -Create copy of SF
- -Send file handle of SF to F1

4025

FIG. 41

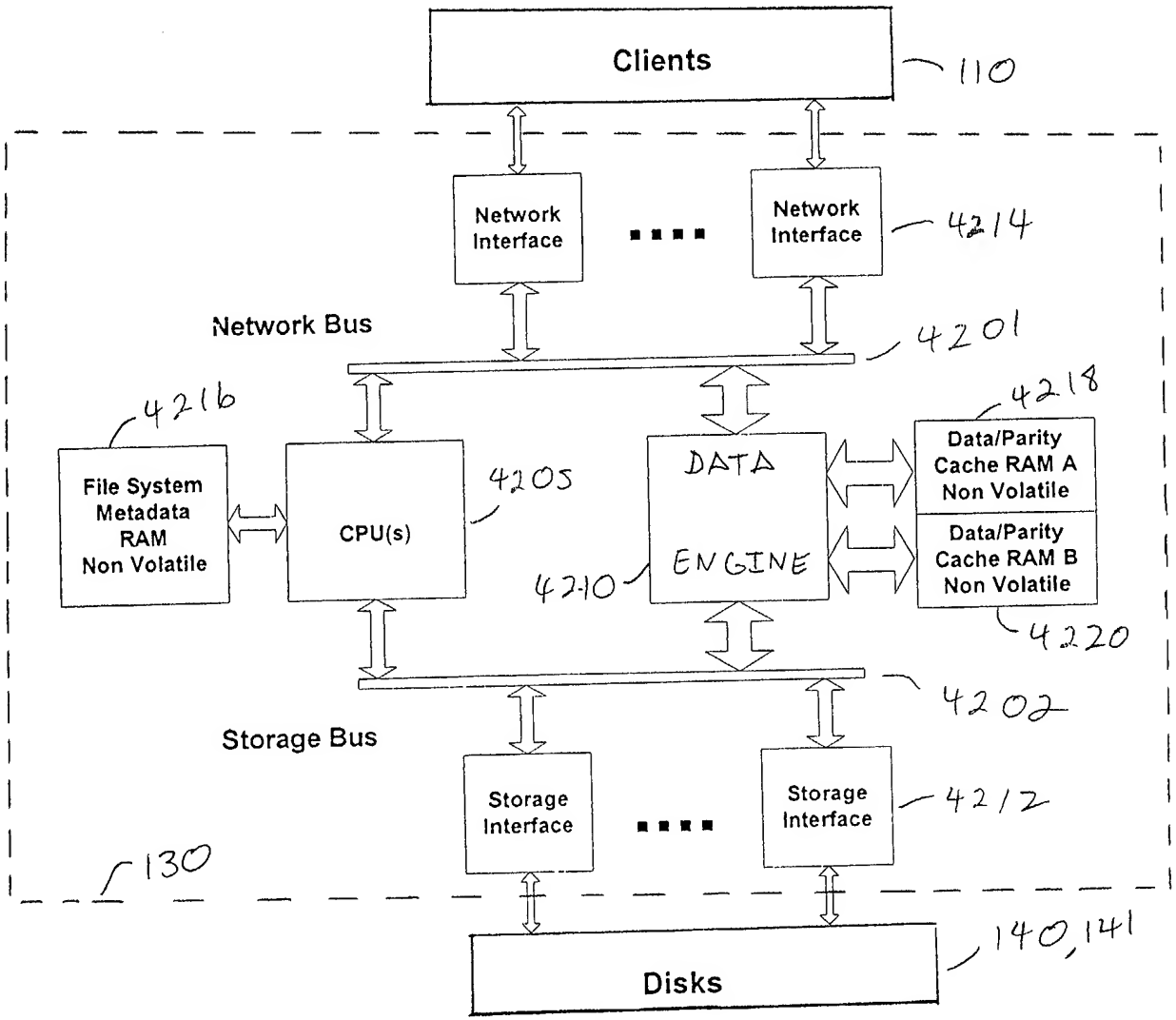

Clients
110
Network Interface
Network Interface
4214
Network Bus
4201
4216
4218
File System Metadata RAM Non Volatile
CPU(s)
4205
DATA
ENGINE
4210
Data/Parity Cache RAM A Non Volatile
Data/Parity Cache RAM B Non Volatile
4220
4202
Storage Bus
Storage Interface
Storage Interface
4212
130
140, 141
Disks

FIGURE 42

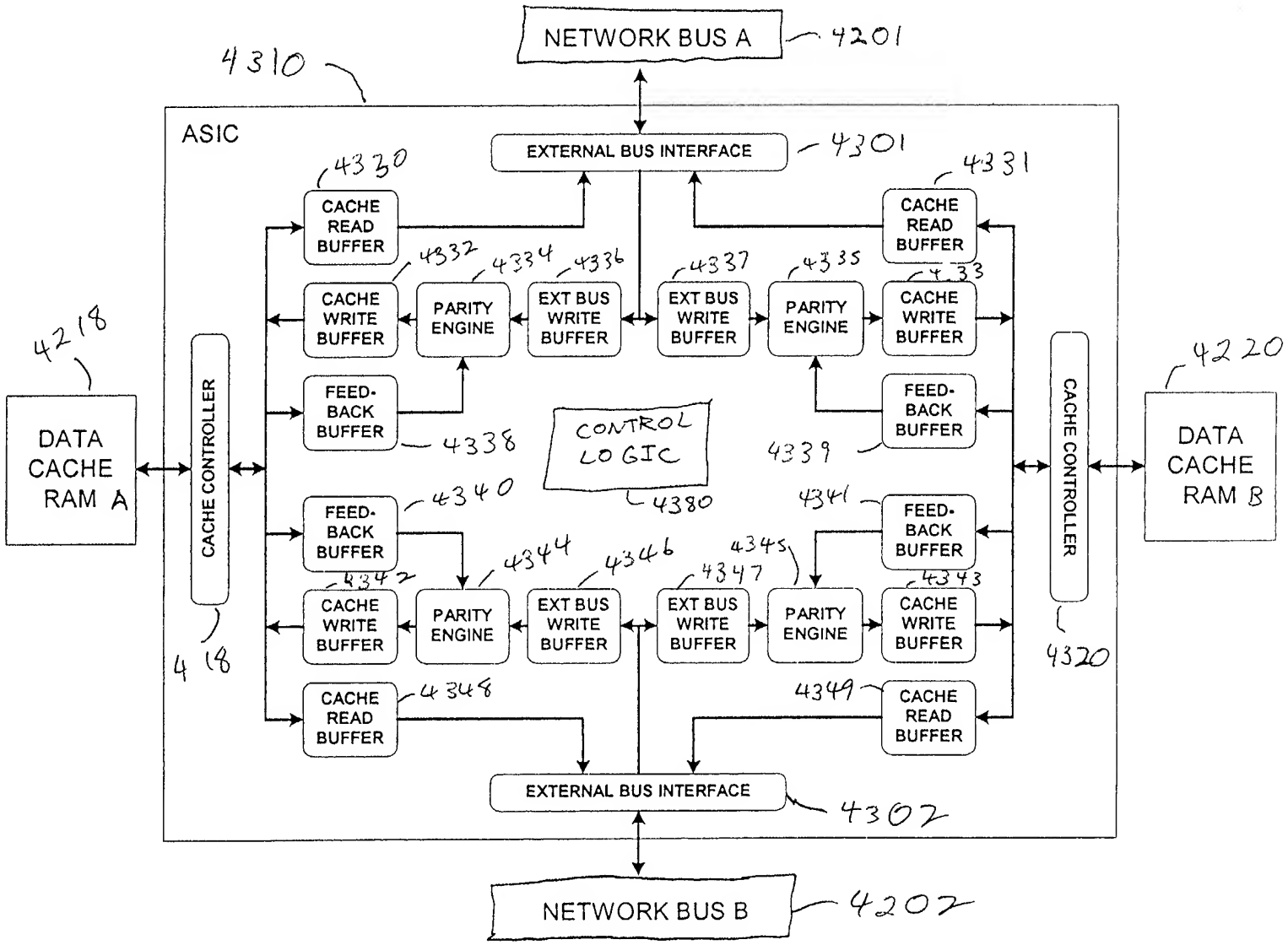
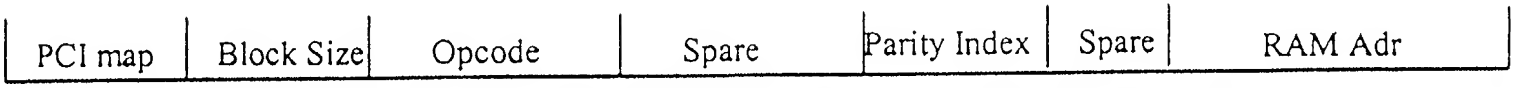

NETWORK BUS A
4201
4310
ASIC
EXTERNAL BUS INTERFACE
4301
4330
4331
CACHE READ BUFFER
CACHE WRITE BUFFER
PARITY ENGINE
EXT BUS WRITE BUFFER
FEED-BACK BUFFER
4332
4334
4336
4337
4335
4333
4338
4339
CONTROL LOGIC
4380
4218
DATA CACHE RAM A
CACHE CONTROLLER
4220
DATA CACHE RAM B
4340
4341
4344
4346
4345
4347
4343
4342
4348
4349
4318
4320
EXTERNAL BUS INTERFACE
4302
NETWORK BUS B
4202

FIGURE 43



4400

FIGURE 44